Exploring the effects of GIS use on students' achievement in geography

Von der Pädagogischen Hochschule Heidelberg
zur Erlangung des Grades einer
Doktorin der Philosophie (Dr. phil.)
genehmigte Dissertation von
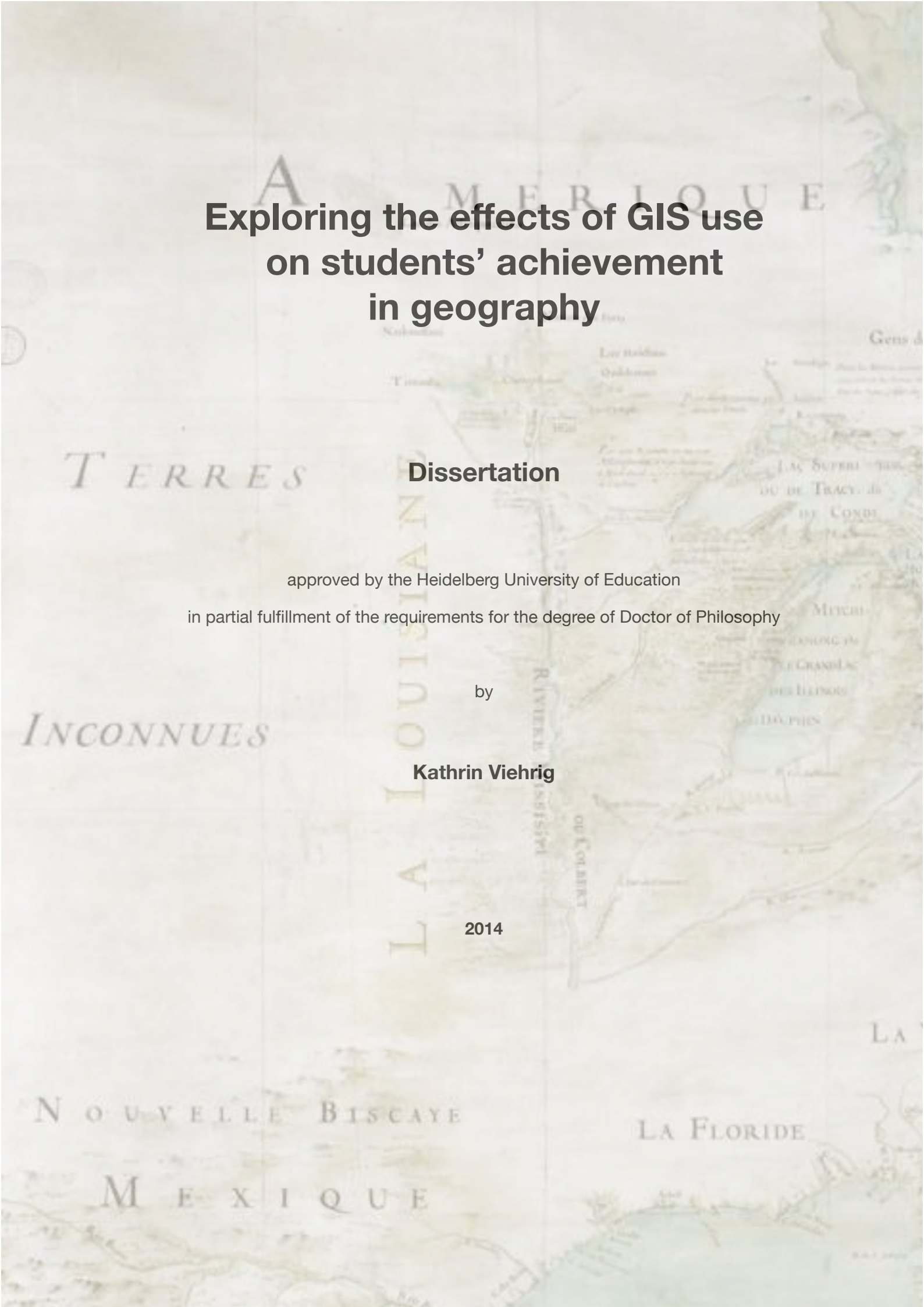
Kathrin Viehrig

aus

Pirna (in Sachsen)

2014

Erstgutachter: Prof. Dr. Alexander Siegmund

Zweitgutachter: Prof. Dr. Nir Orion

Fach: Geographie

Tag der Mündlichen Prüfung: 05.12.2014

Exploring the effects of GIS use on students' achievement in geography

Dissertation

approved by the Heidelberg University of Education

in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Kathrin Viehrig

2014

First advisor: **Prof. Dr. Alexander Siegmund**

Professor of Physical Geography and its Didactics

Head of Department of Geography

Heidelberg University of Education, Heidelberg, Germany

Second advisor: **Prof. Dr. Nir Orion**

Head of Earth and Environmental Sciences Group

Weizmann Institute of Science, Rehovot, Israel

Other committee members:

Prof. Dr. Wolfgang Knörzer

Professor of Sport Science/ Sport Education

Dean of Faculty of Natural and Social Sciences

Heidelberg University of Education, Heidelberg, Germany

Prof. Dr. Manuela Welzel-Breuer

Professor of Physics and its Didactics

Heidelberg University of Education, Heidelberg, Germany

Prof. Dr. Ulrich Michel

Professor of Geoinformatics and its Didactics

Heidelberg University of Education, Heidelberg, Germany

Subject: Geography

Day of the oral examination: 05.12.2014

Acknowledgements

There are many people who provided valuable contributions and support to this dissertation.

First of all, I would like to thank my main advisor Prof. Dr. Alexander Siegmund for his continuing support and provision of learning opportunities as well as helpful discussions. Secondly, I want to thank Prof. Dr. Nir Orion for agreeing to be my second advisor, welcoming me for visits when I was in Israel and very interesting discussions about systemic thinking. Thanks also to Dean Prof. Dr. Wolfgang Knörzer for heading the committee as well as Prof. Dr. Manuela Welzel-Breuer and Prof. Dr. Ulrich Michel for agreeing to be members of the committee.

Many people have helped shape my understanding and skills in geography education, GIS and statistics, have provided feedback, help and new insights, and have proofread draft versions.

First of all, I want to thank the current and former members of the ʳgeo-team at the Department of Geography, Heidelberg University of Education, especially Dipl.-Geoökol. Daniel Volz, Dr. Christina Fiene, Dr. Isabelle Kollar, Dr. Raimund Ditter, Dr. Sebastian Günthert, Dipl.-Geog. Markus Jahn, Dr. Kerstin Voß, Dr. Alexandra Siegmund, Prof. Dr. Peter Dippon, Dr. Nils Wolf and Dipl.-Geog. Svenja Brockmüller. I also want to thank the HEIGIS-members, Dr. Sascha Wüstenberg, Dr. Samuel Greiff and Prof. Dr. Joachim Funke. Additionally, there are many people through which I was inspired or received valuable ideas for this dissertation, such as Prof. Dr. Priyadarsi Roy, Prof. Dr. Kristina Reiss, Dr. Joseph Kerski, Prof. Dr. Sarah Bednarz and many people I talked to at conferences, both nationally and internationally, and courses. Moreover, colleagues from other departments at the University of Education such as Saskia Gensow also provided valuable input, especially at the beginning of the dissertation.

I also want to thank the students and teachers that participated and who thus made this dissertation possible. Thanks as well to the responsible authorities for the permissions. Additionally, thanks also to Dr. Ulrike Klein helping to make the Kiel professional development course happen.

Table of contents

Table of figures

All figures have been created by the author and, in case of empirical results, are based on the studies conducted for this dissertation, unless otherwise noted in the text.

Table of tables

All tables have been created by the author and, in case of empirical results, are based on the studies conducted for this dissertation, unless otherwise noted in the text.

Abbreviations

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AP	Advanced Placement
bin.	Binary
CEFR-L	Common European Framework of Reference for Languages
CFA	Confirmatory Factor Analysis
ch.	Chapter
CTT	Classical Test Theory
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
DGfG	Deutsche Gesellschaft für Geographie (German Association for Geography)
EAP/PV	Expected A Posteriori/Plausible Values
FA	Factor Analysis
Fig.	Figure
GIS	Geographic Information System
GMC	Geographic Methods Competency
GPS	Global Positioning System
GSC	Geographic System Competency
HEIGIS	Heidelberg Inventory of Geographic System Competency
HLM	Hierarchical Linear Modeling
IB	International Baccalaureate
IRT	Item Response Theory
KMK	Kultusministerkonferenz
KW	Kruskall-Wallis
MW	Mann-Whitney
LC	Latent Class
LfS	Landesinstitut für Schulentwicklung (State Institute for School Development)
M	Mean
Mdn	Median

MNSQ	Mean Square
n.s.	Not Significant
PCA	Principal Component Analysis
PISA	Programme for International Student Assessment
RMPE	Recommended Minimum Practically Significant Effect Size
SD	Standard Deviation
SEM	Structural Equation Modeling
Tab.	Table
TIMSS	Trends in International Mathematics and Science Study
W	Wilcoxon-test
WLE	Weighted Likelihood Estimates

To shorten the long names of the different federal state ministries responsible for curricula in the text, they are abbreviated by using MK plus the short form of the respective federal state, e.g. "MKBW" stands for "Ministerium für Kultus, Jugend und Sport Baden-Württemberg".

BB	Brandenburg	NS	Lower Saxony
BL	Berlin	NW	North Rhine-Westphalia
BR	Bremen	RP	Rhineland-Palatinate
BW	Baden-Wurttemberg	SA	Saxony-Anhalt
BY	Bavaria	SH	Schleswig-Holstein
HE	Hesse	SL	Saarland
HH	Hamburg	SN	Saxony
MV	Mecklenburg-Vorpommern	TH	Thuringia

Traditionally, students in the German school system have been streamed into different secondary schools, the high stream (*Gymnasium*), the mid stream (*Realschule*) and the low stream (*Hauptschule*). Numerous special forms exist in different federal states (e.g. the Saxon *Mittelschule* combining mid and low stream students). A more comprehensive overview can be found e.g. in KMK (2014).

Abstract

Technologies based on Geographic Information System (GIS) are widely used in society and are increasingly being integrated into school curricula and practice. Many claims have been made that the use of GIS in class has positive effects on a wide range of achievement and affective variables. However, empirical evidence for that, especially in the German situation, has been scarce.

Systemic thinking has been central to the guiding objective of German geography education for many years and constitutes an important contribution to prepare students for life in a complex world. Yet, so far, specific test instruments and studies elucidating factors that help students improve this competency have been far from extensive.

This dissertation aims at exploring the influence of a short 'working with GIS' vs. 'working with maps' unit on students' achievement in geography, specifically, the systemic thinking competency. Based on literature a definition of geographic systemic thinking and an associated competency model were developed. In total, three one test time and two pre-/posttest with control group studies were conducted to develop test instruments and a treatment as well as to study the question at hand. The treatment used the topic 'tourism in Kenya'. Partly Desktop-, partly Web-based GIS versions were used. In study 5, there were two different types of materials, which contained parallel contents/tasks. While one used an overview sheet of relevant GIS functionality ('old'), the other integrated more step-by-step instruction directly into the text ('new').

Variables included were systemic thinking, sex, age, stream/type of geography study/pre-score, grade/semester, language and migration background, pre-experience, affective variables, pre-basic spatial thinking skills, treatment and material type. Not all variables were used in every study.

The largest study (study 5) used the results of 932 seventh grade students for analyses. The sample contains both high and middle stream students from three German federal states. The study highlights issues such as e.g. test time constraints, open task coding, partly ceiling effects and item difficulties partly deviant from the model expectations. For the analyses, both raw average

scores and WLE estimates obtained by a Rasch analysis are used. Additionally, based on the WLE scores, HLMs are calculated.

Overall, in study 5 GIS students do not improve pre- to posttest in systemic thinking. Consequently, GIS has no positive, and partly a significantly negative impact compared to maps, e.g. in a HLM with all other variables having significant effects included. Results for material type are mixed. For instance, on the one hand, t-tests show no significant difference in pre-posttest-change between students working with 'old' and 'new' WebGIS materials. On the other hand, the overall HLMs with other variables included show a significant negative effect only for the 'old' but not for the 'new' WebGIS materials.

Only 23 students could be included in the 'having already worked with an educational GIS'-sub-group (vs. 520). The improvement of these students pre- to posttest is not significant, but has an effect size above 0.2. A calculation with the 'no pre-experience' sub-group being reduced to students with similar characteristics (e.g. in terms of stream, GIS type) leads to 19 vs. 84 students and similar results. In both cases, students with pre-experience perform not significantly, but with an effect size above 0.2, better than students without pre-experience. Overall, the results could hint at students needing more pre-experience so as to not have so much mental capacity tied to getting used to the software and being able to concentrate more on the system interrelationships. However, due to the sample characteristics and study design, this cannot be proven by the present data and thus needs to be explored in further studies.

Other variables (age, sex, migration and language background, stream, prescore, pre-spatial thinking score) show mixed results depending on the analysis method used. This underlines the impact of methodological choices and the need for large sample studies in order to be able to take a closer look at individual sub-groups. Furthermore, the HLM results point to not all influencing variables having been included. In general, the impact of variables such as pre-achievement/ stream and sex on pre-posttest change evident in some of the analyses points to the need for more research to develop differentiated learning materials.

The conducted studies also show, e.g. through deviations from the assumed model of systemic thinking, that there is still a great need for more studies in terms of test- and model development for systemic and spatial thinking in a geographic context.

Zusammenfassung

Technologien, die auf Geographischen Informationssystemen (GIS) basieren, werden weithin in der Gesellschaft benutzt und zunehmend in Lehrpläne und schulische Praxis integriert. Viele Behauptungen wurden aufgestellt, dass der Einsatz von GIS im Unterricht positive Auswirkungen auf eine große Auswahl von Leistungs- und affektiven Variablen hat. Jedoch sind empirische Belege dafür rar, besonders in der deutschen Situation.

Systemisches Denken ist seit vielen Jahren zentral für das Leitziel des deutschen Geographieunterrichts und stellt einen wichtigen Beitrag dar, um SchülerInnen für das Leben in einer komplexen Welt vorzubereiten. Trotzdem sind bisher spezifische Testinstrumente und Studien, welche Faktoren aufklären, die SchülerInnen helfen, diese Kompetenz zu verbessern, weit davon entfernt umfassend zu sein.

Diese Dissertation hat das Ziel der Exploration des Einflusses einer kurzen 'Arbeiten mit GIS' vs. 'Arbeiten mit Karten'-Unterrichtseinheit auf die Schülerleistung in Geographie, spezifisch der Kompetenz zum systemischen Denken. Basierend auf Literatur wurden eine Definition geographisch-systemischem Denkens und ein damit verbundenes Kompetenzmodell entwickelt. Insgesamt wurden drei Studien mit einmaligem Testzeitpunkt und zwei Prä-/Posttest-Studien mit Kontrollgruppe durchgeführt, um die Testinstrumente und das Treatment zu entwickeln sowie die Fragestellung zu untersuchen. Das Treatment verwendete das Thema 'Tourismus in Kenia'. Teilweise wurden Desktop-, teilweise Web-basierte GIS-Versionen verwendet. In Studie 5 gab es zwei unterschiedliche Materialarten, welche parallele Inhalte/ Aufgaben enthielten. Während die eine ein Überblicksblatt über die relevanten GIS-Funktionen ('alt') verwendete, wurden bei der anderen Schritt-für-Schritt Anleitungen direkt in den Text integriert ('neu').

Einbezogene Variablen waren systemisches Denken, Geschlecht, Alter, Schulart/Art des Geographiestudiums/Prätest-Ergebnis, Klassenstufe/Semester, Sprach- und Migrationshintergrund, Vorerfahrung, affektive Variablen, Prä-

test-Ergebnis im räumlichen Denken, Art des Treatments und der Materialien. Nicht alle Variablen wurden in jeder Studie verwendet.

Die größte Studie (Studie 5) verwendete die Ergebnisse von 932 Siebtklässler-Innen für die Auswertungen. Die Stichprobe enthält sowohl Gymnasial- als auch RealschulschülerInnen aus drei deutschen Bundesländern. Die Studie zeigt Probleme auf, wie z.B. Testzeit-Beschränkungen, Kodierung offener Aufgaben, teilweise Deckeneffekte und Item-Schwierigkeiten, die teilweise von den Model-Erwartungen abweichen. Für die Auswertungen werden sowohl Durchschnitts-Rohwerte als auch WLE-Werte, die durch eine Rasch-Analyse gewonnen wurden, verwendet. Zusätzlich werden, auf Grundlage der WLE-Werte, HLMs berechnet.

Insgesamt verbessern sich GIS SchülerInnen nicht im systemischen Denken vom Prä- zum Posttest. Folglich hat GIS keine positive, und teilweise eine signifikant negative Auswirkung im Vergleich zu Karten, z.B. in einer HLM bei der alle anderen Variablen, die einen signifikanten Einfluss haben, eingeschlossen wurden. Ergebnisse für die Art der Materialien sind gemischt. Zum Beispiel zeigen auf der einen Seite t-Tests keinen signifikanten Unterschied in der Prä-Posttest-Veränderung zwischen SchülerInnen, die mit den 'alten' und die mit den 'neuen' WebGIS Materialien arbeiten. Auf der anderen Seite zeigen die Gesamt-HLMs mit anderen eingeschlossenen Variablen einen signifikant negativen Effekt nur für die 'alten' aber nicht für die 'neuen' WebGIS Materialien.

Nur 23 SchülerInnen konnten in die 'haben schon einmal mit einem didaktischen GIS gearbeitet'-Teilgruppe eingeschlossen werden (vs. 520). Die Verbesserung dieser SchülerInnen vom Prä- zum Posttest ist nicht signifikant, hat aber eine Effektstärke über 0,2. Eine Berechnung, bei der die 'Teilgruppe ohne Vorerfahrung' auf SchülerInnen mit ähnlichen Eigenschaften (z.B. in Bezug auf Schulart, GIS-Art) reduziert wurde, führt zu 19 vs. 84 SchülerInnen und ähnlichen Ergebnissen. In beiden Fällen schneiden die SchülerInnen mit Vorerfahrung nicht signifikant, aber ebenfalls mit einer Effektstärke über 0,2, besser ab als die SchülerInnen ohne Vorerfahrungen. Insgesamt klingt in den Ergebnissen an, dass SchülerInnen mehr Vorerfahrung benötigen, um nicht so viel mentale Kapazität an das Gewöhnen an die Software gebunden zu haben und sich

mehr auf die System-Zusammenhänge konzentrieren zu können. Aufgrund der Stichproben-Eigenschaften und des Untersuchungsdesigns kann dies durch die vorhandenen Daten jedoch nicht bewiesen werden und muss daher in zukünftigen Studien untersucht werden.

Andere Variablen (Alter, Geschlecht, Migrations- und Sprachhintergrund, Schulart, Prätest-Ergebnis, Prätest-Ergebnis im räumlichen Denken) zeigen gemischte Ergebnisse in Abhängigkeit von der verwendeten Analysemethode. Dies unterstreicht die Bedeutung der methodischen Entscheidungen und den Bedarf an Studien mit großer Stichprobengröße, um individuelle Teilgruppen genauer betrachten zu können. Darüber hinaus weisen die HLM Ergebnisse darauf hin, dass nicht alle beeinflussenden Variablen eingeschlossen wurden. Im Allgemeinen weist der Einfluss von Variablen wie Prä-Leistung/Schulart und Geschlecht auf die Veränderung von Prä- zu Posttest, welcher sich in einigen der Analysen gezeigt hat, auf den Bedarf an mehr Forschungsarbeiten hin, um differenzierte Lernmaterialien zu entwickeln.

Die durchgeführten Studien zeigen auch, z.B. durch Abweichungen von den angenommenen Modellen systemischen Denkens, dass immer noch ein großer Bedarf an mehr Studien in Bezug auf die Test- und Modellentwicklung für systemisches und räumliches Denken in einem geographischen Kontext besteht.

1 Setting the scene

"Now more than ever, we need people who think broadly and who understand systems, connections, patterns, and root causes, how to think in whole systems, how to find connections, how to ask big questions, and how to separate the trivial from the important."
Orr (1994), in Kerski (2009)

"We need to be clear about the role of GIS in promoting the core tenets of geography before proceeding to invest continued time and effort on its behalf. It is not sufficient for us to assume that GIS promotes and develops spatial skills. We need to know if it occurs and under what conditions in order to further expand its practice - or to find better and simpler ways to achieve the same goals."
Bednarz (2001, p. 4)

These two quotes summarize the rationale for this dissertation.

1.1 GIS didactics: Educational white lands

In the past, in between colorfully mapped areas, maps often showed vast white lands (see cover image). *Terres inconnues* – areas that for many people were lands not yet explored, documented and shared. Despite considerable progress in the last years, the area of GIS didactics can still be represented by a map with considerable white lands.

GIS didactics covers everything relevant to the science and art of teaching and learning with and about GIS (Geographic Information Systems) – for instance content, methods, objectives, as well as persons and their relationships (cp. e.g. for GIS education Baker, Kerski, Huynh, Viehrig, & Bednarz, 2012; for an overview of general definition of didactics Petersen, 2007, pp. 19-21; Uljens, 1997, pp. 44-46). Thereby,

"[a] geographic information system is a combination of elements designed to store, retrieve, manipulate, and display geographic data – information about places. It is a package consisting of four basic parts: robust hardware, powerful software, special data, and a thinking explorer." (ESRI, 1998, p. 2) (see also e.g. de Lange, 2005; Falk, 2004; Schallhorn, 2004b).

A GIS normally contains both layers of graphics (e.g. maps of rivers, farms, ...) and corresponding attributes stored in a table, although the graphic display of the spatial co-ordinates is not mandatory (cp. e.g. ESRI, 1998; Naumann, Volz, & Viehrig, 2008). Thereby, a GIS is a "[...] model of the real world [...], which enables different graphical as well as professional views on a dataset" (de Lange, 2007, p. 35, translated).

Over the past decades, GIS didactics has become an area of global interest. The GIS software company ESRI has users in more than 150 countries (ESRI, n.d.) – including almost every major university (Phoenix, 2004). GIS education programs already at school level exist in many countries, for instance "[...] Australia, Canada, Denmark, Germany, Sweden, Switzerland, and the US" (Phoenix, 2004, p. 2), Jamaica, Finland, Nepal, New Zealand, Rwanda, and Taiwan (cp. e.g. Brodie, 2004; Forster, 2008; Kerski, 2008b; Ministry of Education Youth and Culture Jamaica, 2004; Nepal Geographic Information System Society, 2008; Olsen, 2001; Tolvanen, 2008). In a recent book on GIS in secondary schools, authors from 33 countries are represented (Milson, Demirci, & Kerski, 2012). The "ability to create thematic maps with GIS" has also been included in the survey by Hemmer, Hemmer, Obermaier and Uphues (2008, p. 27, translated, n=282), receiving a perceived importance average score of 2.38 from society representatives and 2.63 from geography experts, i.e. on average 2.48 (out of 5). However, this was the lowest score of all 41 items.

Despite global interest and a growing body of literature, there still are large gaps in the knowledge about the area of GIS didactics. Consequently, in publications, conferences (e.g. the meetings of the GIS education research work group 2008) and private conversations a call for 'explorers' conducting further research in all areas of GIS didactics and 'mapmakers' synthesizing and ordering the existing evidence into a coherent theory of learning and teaching with and about GIS has been issued again and again (see e.g. Baker, et al., 2012). The lack of such a theory and its foundational empirical research can be perceived as one of the impediments to a spread of GIS implementation in education (e.g. Baker & Bednarz, 2003; Baker, et al., 2012; Siegmund, Volz, & Viehrig, 2007) which goes beyond constraints in areas such as time, hard- and software suitability, support, availability of materials, training characteristics and

perceived potential explored in a number of teacher surveys (e.g. Baker, Palmer, & Kerski, 2009; Höhnle, Schubert, & Uphues, 2010, 2011; Johansson, 2003; Kerski, 2003; Yuda & Itoh, 2007).

1.2 Significance of exploration endeavors

For centuries educators have striven to improve their students' education, evaluating a large number of proposed improvements in pedagogy, content, methods, media etc.. Any proposed change, however, needs to be seen as one part integrated into a complex whole. Many models seek to analyze the complexity of education and identify important elements and interactions, for instance the offer-use-model by Helmke (2003), the Berlin model developed by Heimann (see Peterßen, 2000, pp. 82-95) or the "[...] interactions that occur during episodes of guided learning" described by Branch and Gustafson (1998, p. 4). GIS and other media are thus only a small part of the educational experience.

Despite this, media have received considerable attention in the last years. They carry with them a multitude of promises for new ways of helping students learn (see ch. 2.2.3). Moreover, in a world where information is increasingly available and where autonomous, life long learning has become pivotal, being able to use media competently has become a basic skill. This is also true for GIS and other spatial applications, leading for instance to a discussion of digital geo-media in the context of an education for spatial citizenship (e.g. HERODOT, 2008) in the wake of developments such as GPS-enabled smart phones or the INSPIRE initiative "[...] to create a European Union (EU) spatial data infrastructure" (European Commission, n.d.).

Yet, the impact of any one media choice versus another as part of the educational process is controversially discussed. One webpage (Russell, 2010) lists dozens of comparative media studies, especially between face-to-face and distance (e.g. online) delivery, with many showing no significant difference. Moreover, many media innovations such as laptop classes, which had been lauded at the beginning, are being discontinued (e.g. Popp, 2007). In general, studies such as Baker & White (2003) or Bednarz & Bednarz (2008) attest to considerable instructor effects. Moreover, research also shows differences in the effects on various kinds of pupils (e.g. Kerski, 2000). Improving student

achievement seems to thus be primarily a matter of people – yet, it could be argued that media can hinder or help learning and teaching to some extent.

A number of articles contain enthusiastic propositions regarding what a positive difference GIS could make to students' education. However, there is a large gap between these propositions and the number of studies which actually give evidence to what Bednarz describes as "[...] the role of GIS in promoting the core tenets of geography [...]" (2001, p. 4). Especially in the German context, only very few studies examine the effects of GIS use in school education. However, "[i]t is not sufficient for us to assume that GIS promotes and develops spatial skills. We need to know if it occurs and under what conditions in order to further expand its practice - or to find better and simpler ways to achieve the same goals" (Bednarz, 2001, p. 4). Or, as Briebach (2007, p. 10) argued in light of the available body of research at that time: "Though encouraging initial research, when deciding whether or not to invest in GIS, this is hardly compelling evidence of its impact."

To test proposed options for the improvement of spatial skills and other 'core tenets' of geography is important, since "[...] many contemporary situations that citizens encounter have a base in "Earth systems"[...]" (OECD, 2007b, p. 114) (see also e.g. Bednarz, Heffron, & Huynh, 2013). Published estimates state that "[...] approximately 80 percent of all decisions in the public and private life bear a spatial reference" (Bundesamt für Kartographie und Geodäsie, 2003, p. 8, translated).

Being good in geography (and earth science) is not easy, however. Both German (e.g. Hüttermann, 2004) and international (e.g. Beaton, et al., 1996; Martin, et al., 2008; National Geographic Education Foundation, 2002; Niemz & Stoltman, 1993; OECD, 2007c) studies have shown that from primary school to university/ young adult level, people in Germany often have difficulties with various aspects of the geography or earth science domains (for the situation in the USA see e.g. Bednarz, et al., 2013; Roper Public Affairs & National Geographic Education Foundation, 2006; for an overview of countries participating in some international assessments see e.g. Siegmund & Viehrig, 2012). For instance, in the PISA 2006 survey students in Germany had a mean score of 516 for the 'physical systems'

but of only 510 for the 'earth and space system' scale (OECD, 2007c). Thereby, for instance both the InterGEO II study and the Geographic Literacy Survey showed quite some variation between individual items and areas (National Geographic Education Foundation, 2002; Niemz & Stoltman, 1993). In the InterGEO II international comparative study of geographic achievement, German 14-year-old students overall had an average of only 61.3% correct answers. By sub-scale, German students scored highest on I (Location, 77.1%) and III (Human Geography, 70.9%), lower on VI (Geography of the home country, 61.1%) and IV (Geographical Skills 60.0%), and lowest on V (Regional Geography, 55.2%) and II (Physical Geography, 50.6%) (Niemz & Stoltman, 1993, p. 5).

Apart from the pedagogical rationale, there is an economic one. What are cost-efficient ways to improve education? Levin (2001, p. 55) states that "[...] no one has argued that educational spending is highly efficient". In 2006, the German state paid 50.8 billion euros for the primary and secondary schools alone, not to speak of additional financing through the private sector (Statistisches Bundesamt, 2009). Beyond the monetary costs involved, however, there are costs in time. Students generally spend at least nine to twelve years of their limited life time going to school. If the quality of the outcome is poor, the resulting costs, both for the individual and society at large, are very high (cp. e.g. Bertelsmann Stiftung, n.d.; BMBF, 2008; Weinert, 2001).

Additionally, the costs need to also be taken into account when considering new media such as GIS. While GIS integration is said to have multiple benefits for the various areas of the educational process, outcome and efficiency (see ch. 2.2.3), it is not itself without costs (see ch. 2.2.4). Yet, exactly these costs also point to another educational economic factor to consider: employment opportunities. GIS training has only been recently integrated into some pre-service teacher programs. Training teachers as well as providing GIS education consulting, materials, software and curriculum relevant data thus constitutes a global job opportunity.

Beyond school education, geospatial technology such as GIS is seen as an important current industry field (see e.g. Baker, et al., 2012; Kerski, 2008a; Schleicher, 2004a). To estimate the global actual use of GIS and GIS-based applications is difficult, especially due to the wide availability of free online products. At

the beginning of the millennium, there were an estimated 140,000 organizations worldwide that used GIS (Gewin, 2004). In 2009, the annual ESRI user conference alone attracted, despite the economic crisis, "more than 12,000 attendees from 104 countries" (Wheeler, 2009). Jobs requiring GIS-skills are spread across a wide variety of industries – city planning, military defense, environmental protection, utilities and communications, customer analysis, archaeology and emergency responses, to name a few. Consequently, the demand for GIS specialists has been growing worldwide (Gewin, 2004; Kerski, 2008a), especially for people who have completed more than basic training (Fitzpatrick & Dailey, 2000). An estimated one million people worked with GIS as part of their job around the turn of the millennium, their number growing by approximately 15% annually (Fitzpatrick & Dailey, 2000; Glover, 2006). Thereby, demand has been greater than supply (Fitzpatrick & Dailey, 2000; Glover, 2006). Consequently, integrating basic GIS training already into school education has been proposed as one way to increase the supply and providing students with job relevant skills (e.g. Cremer, Richter, & Schäfer, 2004; Falk & Schleicher, 2005).

Besides their role in the working world, media such as GIS have become an important part of the lives of young people and society at large. The JIM study showed that 89% of the sampled twelve to 19-year-olds in Germany used a computer and 84% the internet several times a week and 95% owned a cell phone (Medienpädagogischer Forschungsverbund Südwest, 2008) (for more information regarding computer/internet use in and outside of the classroom see also e.g. European Commission, 2006; Herzig & Grafe, 2006; Medienpädagogischer Forschungsverbund Südwest, 2012; OECD, 2011; Pfeiffer & Pfeiffer, 2011; Rauh, 2006; Rumpf, Meyer, Kreuzer, & John, 2011). All these platforms offer a variety of opportunities to come in contact with geospatial technology-based applications, for example cell phone navigation, route planners, branch finders, digital city maps and information services of government agencies. Consequently, being able to use geospatial technologies could be considered part of reaching the overall educational aim of "[...] ensuring that all adolescents of a generation, independent of background and sex, are enabled to live [...] in autonomous participation in politics, society and culture, and in the

shaping of the own lifeworld, and self-determinedly act as responsible citizens"
(Klieme, et al., 2007, p. 63, translated).

GIS and other geospatial technologies have been explicitly integrated into the
secondary school curricula of a number of German federal states, especially
with regard to the high stream (e.g. Siegmund & Naumann, 2009; Viehrig &
Siegmund, 2012). Moreover, it has been included into the German national
educational standards for geography (DGfG, 2007a). GIS has also been named
as example for the inquiry project presentation in the common demands for the
Abitur, the school leaving certificate usually obtained at the end of the twelfth
grade which is necessary to study at a university (KMK, 2005b). Moreover, it
has been integrated into the framework for geography teacher education
(DGfG, 2010; KMK, 2008). Even where it is not explicitly included, however, it
can be seen as part of areas such as 'using modern information technologies'
or 'methods competency' (see e.g. Viehrig & Siegmund, 2012). Internationally,
GIS and other geospatial technology has been integrated into, for instance, the
AP standards for human geography (College Board, 2007) in the USA or the
International Geographical Union's declaration on Geographical Education for
Sustainable Development (Haubrich, Reinfried, & Schleicher, 2007).

In general, changes in curricula and other parts of the educational framework
are not always well received. The societal and professional discourse contains
a multitude of complaints about the structure/ structural changes of the educa-
tional system, the competencies of graduates and the content of the curricula
(cp. e.g. cpa/dpa/AP, 2006; mer/dpa, 2007; Philologenverband Baden-
Württemberg, 2008; StN, 2010). For instance, a study in the federal state
Baden-Württemberg showed that in each of the different school types, less
than 5% of the sampled teachers (*n*=1140) thought that the changed curricu-
lum instituted in 2004 is a project well done (GEW, 2004).

1.3 Research objective

GIS didactics – and this dissertation – are thus placed on the intersection of
multiple contexts. This dissertation concentrates on exploring the educational
potential of GIS for one of the "core tenets of geography" (Bednarz, 2001, p. 4).
– selected aspects of geographic system competency – within part of the Ger-

man secondary school system. Does GIS use, in the way it is currently often practiced, really help improve the competency achievement of students? Is it worth it? What are some of the conditions under which GIS use has a positive or negative effect? To the answers to these pivotal questions this dissertation wants to be provide a small contribution.

To address these questions, chapter 2 will first give a brief overview over some of the pertinent literature, both with regard to geographic system competency and GIS in education. Chapter 3 will describe the methodology, including the research hypotheses. Chapter 4 will briefly describe the achieved samples. Chapter 5 describes the test instruments. Chapter 6 describes the development of the learning unit. Chapter 7 presents the results of data analysis by hypothesis as well as HLM analyses combining different variables. Chapter 8 will discuss key implications and outlooks.

2 Reviewing the literature

A literature review sketches a 'map' of an educational area. Like a map, it can be an important help in the decision-making process of where one could move, what consequences such a move could have, which resources are needed, and which factors might influence the process. Explorers constantly publish their findings and interpretations of existing data may change as a result.

Existing studies in the areas of geographic system competency, of the impact of GIS use and the effects of student variables on this impact differ widely in their objectives, methodology, theoretic foundations and implications as well as task formats, making it difficult to conduct meta-analyses and drawing the implications together into a coherent theory. This is also one of the problems discussed with geography education research in general (e.g. Bednarz, et al., 2013).

2.1 Competencies in geography education

> "Education as process, in subjective meaning, is equipment for behavior in the world." (Robinsohn, 1967, p. 13, translated)

To the overall objective of (school) education, geography education's guiding objective should be a subject specific contribution (see e.g. Hoffmann, 2000; Köck, 1993, 1999). The guiding objective describes the subject's overriding aim. All other objectives, as well as the selection of content, media and methods employed in the lessons throughout the school year are to be geared towards it and help in achieving it (e.g. Haubrich, 2006; Köck, 1997, 2005a). This also includes the use of GIS (Volz, Viehrig, & Siegmund, 2008).

In the German discourse, proposed, and controversially discussed, guiding objectives such as spatial behavior competence *(Raumverhaltenskompetenz),* creation competence for sustainable development (*Gestaltungskompetenz für nachhaltige Entwicklung)* and "[...] insight into the connections [...]" and "[...] space-oriented action competence" *("[...] Einsicht in die Zusammenhänge [...]" und "[...] raumbezogene Handlungskompetenz")* differ in the details of what is expected of the students, but all show that the main goal of geography education has been competency achievement (e.g. DGfG, 2007a, p. 5; Haubrich, 2006; Köck, 1977,

1989, 1993, 2003, 2005b; Lethmate, 2001; Rinschede, 2005; Viehrig & Volz, 2013) (for discussion of other guiding ideas see e.g. Flath, 1994; Flath & Fuchs, 1994; Haubrich, 1982, 2006; Köck, 1989, 1997, 1986; Kross, 1992, 1994; Schallhorn, 2004a; Wilhelmi, 2006). For the situation in the USA, a recently published "roadmap for research" also includes an overview regarding the search for a "[...] common destination [...]" in terms of "[...] learning outcomes [...]" for geography education (Bednarz, et al., 2013, p. 19).

In general, however, 'competency', is a widely used and consequently at times fuzzy term (Erpenbeck & von Rosenstiel, 2003; Hartig & Klieme, 2006; Kaufhold, 2006). The DFG-priority program 1293 "Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes" defines competencies as "[...] context specific cognitive achievement dispositions, which functionally refer to situations and demands in specific domains" (Klieme & Leutner, 2006b, p. 4, translated, in original partly italic). Other definitions additionally include non-cognitive components such as practical skills, motivations, attitudes, emotions, volitions and values (cp. e.g. Einhaus, 2007; Hipkins, 2006; Kaufhold, 2006; Klieme & Leutner, 2006a; Rychen & Salganik, 2003; Schecker & Parchmann, 2006; Weinert, 2001). The context specificity sets competencies apart form general intelligence (Klieme & Leutner, 2006a, p. 879)

While a considerable amount of discussion about the proposed guiding objectives has been generated in the last decades, adequate test instruments and comprehensive studies gauging the students' achievement of the objectives are still largely lacking.

2.1.1 Significance of geographic system competency

A comparison of the proposed guiding objectives shows that analyzing and understanding geographic systems are a central part (Fig. 1). For instance, the national educational standards state that "[t]he main goals of geography lessons are therefore to provide insights into the connections between natural conditions and social activities in different parts of the world, and to teach an associated spatially-oriented competence that can be applied" (DGfG, 2007b, p. 5). Moreover, understanding systems is also part of the European discussion about core competencies of geographic literacy for university education (Don-

ert, 2007). It also plays an important role in international documents, such as a draft for revising the national standards for geography education (NCGE, 2009) and the roadmap for geography education research (Bednarz, et al., 2013) in the USA or a guide to the IB diploma in geography (IBO, 2005). It is also reflected in the objective of geography as a science, defined as "[...] an understanding of the vast, interacting system comprising all humanity and its natural environment on the surface of the earth" (Ackerman, 1963, p. 435). An overview of discussed core elements of geography as a science can be found e.g. in Bednarz et al. (2013).

spatial behavior competence (Köck, 1993)	education for sustainable development (Haubrich, 2006)	"[...] insight into the connections [...]" and "[...] space-oriented action competence" (DGfG, 2007a, p. 5)						
"thinking and acting in spatial ethical categories" (p. 20)		"Knowledge specific to the subject" "Ability to understand spaces at difference scales as physical and human geographical systems and to analyse the interrelations between man and environment" (DGfG, 2007b, p. 9)	"Spatial orientation" (DGfG, 2007b, p. 9)	"Gathering information / methods" (DGfG, 2007b, p. 9)	"Communication" (DGfG, 2007b, p. 9)	"Evaluation" (DGfG, 2007b, p. 9)	"Action" (DGfG, 2007b, p. 9)	
"thinking and acting in space laws and models" (p. 19)								
"thinking and acting in global connections" (p. 19)	"creation competence for a sustainable development [...]" (p. 13)							
"thinking and acting in geo-ecological systems" (p. 18)	(includes "ability for system analysis and problem-solving [...]" as well as "interaction and dynamics perceiving, networked thinking" and "side effect recognizing, forward-looking thinking" (p. 13, with reference to BLK, 1999))							
"thinking and acting in spatial processes" (p. 17)								
"thinking and acting in spatial structures" (p. 17)								
"spatial orientation ability" (p. 16)								

Fig. 1: Structure of selected proposed guiding objectives for geography education (DGfG, 2007a, b; Haubrich, 2006; Köck, 1993) (partly translated)

Systems are a central concept not only in geography. Be it understanding the interrelationships between parts in a cell or understanding the solar system, systemic thinking can be done at different scales and in different contexts. Consequently, systems are called one of the most important cognitive constructs of science (Klaus, 1985; Smithson, Addison, & Atkinson, 2002). Systemic thinking also plays a key role in many everyday decision processes. The term system is included in the current national versions of the educational standards of many subjects, for instance, biology, chemistry, computer science, geography, German language, history, mathematics, physics, politics and technology (DGfG, 2008; GI, 2008; GPJE, 2004; KMK, 2003a, b, 2004a, b,

2005a; VDI, 2007; VGD, 2007). Systemic thinking is also important part of the competencies discussed for global learning (e.g. Rost, 2005b; Schreiber, 2005) To differentiate the competencies needed to deal with *geographic* systems from that needed, for instance, to deal with computer systems, this dissertation will use the term "geographic system competency".

2.1.2 Possible structures of geographic system competency

Competencies can be organized with the help of competency structure models (e.g. Einhaus, 2007; Hemmer & Hemmer, 2007, 2013; Klieme & Leutner, 2006a). A competency structure model consists of dimensions, divided into components (e.g. Einhaus, 2007; Schecker & Parchmann, 2006, cp. Fig. 2). The components of one dimension can be either the "graduation of one skill", "qualitatively different skills" or "parallely developing skills" (Einhaus, 2007, pp. 171-172, translated). In the first case, the components are competency levels (Einhaus, 2007; Schecker & Parchmann, 2006). The development and validation of competency models is a central task for geography and other subjects' didactics (e.g.Hemmer & Hemmer, 2013; Hemmer, 2008a; Klieme & Leutner, 2006a; Viehrig, 2013).

dimensions	components			
dimension 1		component 2	...	component n
dimension 2	component 1	component 2	...	component n
...	component 1	component 2	...	component n
dimension n	component 1	component 2	...	component n
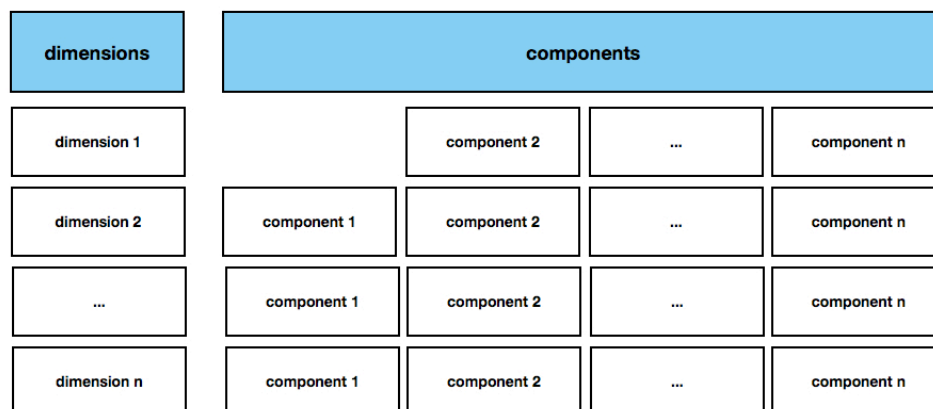
Fig. 2: Composition of competency structure models
      (based on Einhaus, 2007, pp. 170-171)

There is a multitude of theoretical approaches regarding the structure of geographic system competency. For instance, the layers of space-related systems thinking (Köck, 1985), the levels of consideration of space system classes (Köck, 1989), the "thinking and acting in geo-ecological systems" in the framework of spatial behavior competence (Köck, 1993, p. 18, translated), the

levels of ecological thinking (Lecher, 1997), the transfer of Bloom's Taxonomy of cognitive learning targets to geo-ecological systems thinking (Rempfler, 1998), the deliberations about insight-guiding approaches (Köck & Rempfler, 2004), or the division of the area "subject knowledge" of the national educational standards for geography in sub-skills and standards (DGfG, 2008). Recently, these have been supplemented by two theory-guided developments of geography system competency structure models (e.g. Rempfler & Uphues, 2010b; Viehrig, Greiff, Siegmund, & Funke, 2011).

Both in Germany and internationally, there is also a limited number of empirical studies in geoscience contexts elucidating the possible structure of geographic system competency (e.g. Ben-Zvi Assaraf & Orion, 2005a; 2010a; 2010b; Dickerson & Dawkins, 2004; Gudovitch & Orion, 2001; Kali, Orion, & Eylon, 2003; Lücken, Hlawatsch, & Raack, 2005; OECD, 2007b; Orion & Basis, 2008; Raia, 2005; Sibley, et al., 2007; Uphues, 2007).
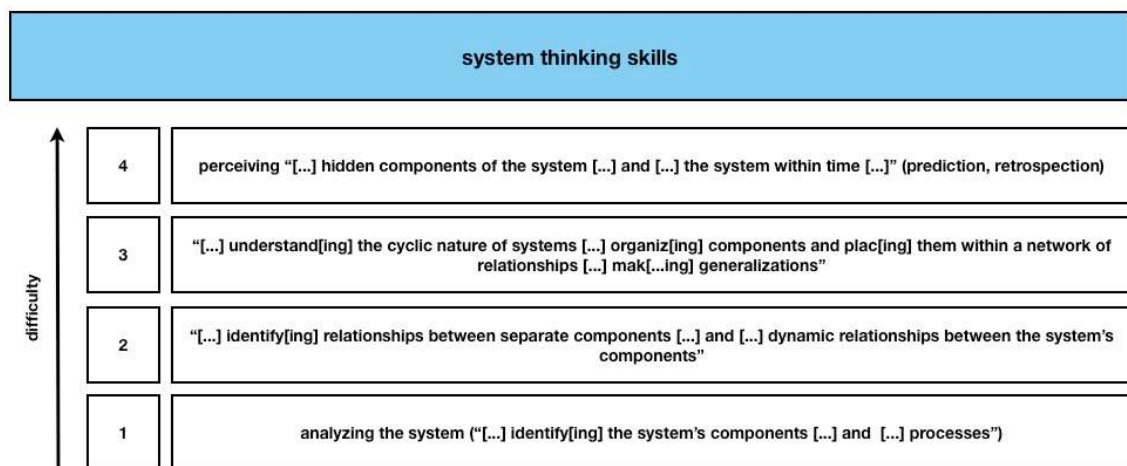
Fig. 3: Example of an empirically derived structure of systemic thinking skills
      (adapted from Ben-Zvi Assaraf & Orion, 2005a, p. 556)

These differ widely in their scope, sample, methodology, generalizability and results. For instance, based on a literature review, Ben-Zvi Assaraf and Orion (2005a) assumed eight distinct systemic thinking skills. An analysis of the data collected with the help of different research tools, some only used with subsamples, showed four hierarchical groups (Fig. 3). In general, students that achieved higher levels also achieved the levels below, i.e. students achieving level four also had the skills of level three, two and one, students achieving level three also had the skills of level two and one etc.. The four levels were

largely confirmed in the study of Orion and Basis (2008), with only the 'perception of hidden elements' appearing lower in the hierarchy. In a later paper, Ben-Zvi Assaraf and Orion (2010a) organized the eight systemic thinking skills into an hierarchical model with three groups, namely "[...] (a) analysis of system components (characteristic 1); (b) synthesis of system components (characteristics 2,3,4,5); and (c) implementation (characteristics 6,7,8)" (p. 3) (see also Ben-Zvi Assaraf & Orion, 2010b).

These levels reflect parts of the definition of a geographic system. Geographic systems consist of elements with their attributes and relationships between these elements from the domains of the geosphere, atmosphere, hydrosphere, biosphere and pedosphere and are more than the sum of their parts (e.g. Ackerman, 1963; Bauer, Englert, Meier, Morgeneyer, & Waldeck, 2002; Haggett, 2004; Hlawatsch, Hildebrandt, Bayrhuber, Hansen, & Thiele, 2005; Köck, 1999; Rempfler, 1999; Rempfler & Uphues, 2011a; Rhode-Jüchtern, 2009; Sommer, 2005; Steinhardt, Blumenstein, & Barsch, 2005; Strahler & Strahler, 2002). Between the parts of the system, energy, matter and information are exchanged (e.g. Ackerman, 1963; Mosimann, 2007; Strahler & Strahler, 2002). The boundaries defining the 'inside' and 'outside' of a system are dependent on the study purpose and hence subjective (e.g. Hlawatsch, Hildebrandt, et al., 2005; Kriz, 2000; Rempfler & Uphues, 2011a; Sommer, 2005; Steinhardt, et al., 2005).

Besides domain specific competency aspects, also domain independent ones have been assumed when dealing with complex systems (cp. e.g. Schecker, Klieme, Niedderer, Ebach, & Gerdes, 1999; Sommer, 2005). Consequently, both theoretical and empirically derived structures of systemic thinking skills in other domains (e.g. Arndt, 2006; Davidz, Nightingale, & Rhodes, 2004; Maierhofer, 2001; Ossimitz, 2000; Rost, Lauströer, & Raack, 2003; Sommer, 2005; Sweeney & Sterman, 2000; Vanasupaa, Rogers, & Chen, 2008) and deliberations regarding the structure of domain unspecific systemic thinking (e.g. Frischknecht-Tobler, Nagel, & Seybold, 2008; Stave & Hopper, 2007) or of the related concept of problem solving (e.g. Greiff & Funke, 2008; Klieme, Funke, Leutner, Reimann, & Wirth, 2001) may give additional insights into the structure of geographic system competency. Frischknecht-Tobler, Kunz and Nagel

(2008), for example, differentiate systemic thinking into system concepts, habits, tools and competencies.

Largely unexplored is what differentiates tasks used to assess systemic thinking in geography from tasks that have been used for instance in biology (e.g. a farmer intervening in an agricultural system, Rieß & Mischo, 2008a), mathematics (e.g. the Hilu tribe, Ossimitz, 2000) or economics (e.g. market price regulation, done within a physics unit, Bell, 2004). It is still an open question whether there are domain specific underlying system competency levels that can be observed across different subjects – including geography. Such competencies could be likened to the CEFR-L (Council of Europe, 2001), which delineates levels of competencies across many different languages. To "[...] study what concepts are unique or especially relevant and appropriate to geography and how to best develop them" is one of the areas the USA geography education research roadmap calls for (Bednarz, et al., 2013, p. 27).

Very recently, the exploration of the structure of geographic system competency has received more attention in geography education research in Germany. For instance, Uphues and Rempfler have developed a theoretical competency model focusing on systemic thinking, whose dimensions are not geography specific and are being studied empirically (Rempfler & Uphues, 2010a, 2011a, b, c, 2012). They delineate the dimensions "[...] 'system organisation', 'system behaviour', 'system-adequate intention to act' and 'system-adequate action'" (Rempfler & Uphues, 2012, p. 11). In contrast, the HEIGIS (Heidelberg Inventory of Geographic System Competency) model contains three dimensions, where the level descriptions for dimension 1 and 2 focus on effect relationships and could fit other domains, but dimension 3 is very geography focused (Fig. 4; for the theoretical development see e.g. Viehrig, et al., 2011; Viehrig, Siegmund, Wüstenberg, Greiff, & Funke, 2012). The project is partly based on this dissertation. The study comprised several empirical studies with university students, leading to model changes (e.g. Funke, Siegmund, Wüstenberg, & Viehrig, 2011; Siegmund, Funke, Viehrig, Wüstenberg, & Greiff, 2012). In both quantitative studies, dimension 2 could not be identified. In the first quantitative study, the remaining two dimensions could be empirically separated, while in the second quantitative study, they could not (Siegmund, et al., 2012).

	Dimension 1: comprehend and analyze systems	Dimension 2: evaluate possibilities to act towards systems	Dimension 3: spatial thinking
level 3	identification and understanding of the complex network of relationships	also take into account side effects and autoregressive processes	can use 8 or more spatial thinking skills
level 2	identify and understand relationships between the system elements	take into account multiple effects	can use 6 or 7 spatial thinking skills
level 1	identify and understand system elements	take into account main effects	can only use up to 5 spatial thinking skills

Fig. 4: HEIGIS model after the CogLabs
(based on Siegmund, et al., 2012, translated)
(largely based on Ben-Zvi Assaraf & Orion, 2005a; Gersmehl & Gersmehl, 2006; Greiff & Funke, 2008 and CogLab results)

2.1.3 Possible achievement of geographic system competency

The great variation in study designs and underlying theoretical models make it difficult to compare achievement in systemic thinking across studies, both inside and outside of geography. Across subjects, however, studies show that systemic thinking poses considerable challenges for students. While studies such as Sommer (2005) or Ben-Zvi Assaraf and Orion (2010b) demonstrate that even primary school children can evidence some basic systems thinking, other studies show that even secondary school and university students still have problems in dealing with system tasks (e.g. Ben-Zvi Assaraf & Orion, 2010a; Bollmann-Zuberbühler, 2008; Cronin, Gonzales, & Sterman, 2009; Falk & Nöthen, 2005; Greiff & Funke, 2008; Gudovitch & Orion, 2001; Hildebrandt, 2006; Hlawatsch, Lücken, Hansen, Fischer, & Bayrhuber, 2005; Kali, et al., 2003; Kaminske, 1996; Orion & Basis, 2008; Ossimitz, Kotzent-Pietschnig, Kreisler, Waiguny, & Zoltan, 2001; Raia, 2005; Sell, Herbert, Stuessy, & Schielack, 2006; Sweeney & Sterman, 2000; Uphues, 2007). For instance, the study by Ben-Zvi Assaraf and Orion (2005a, p. 556) with 70 eighth grade students in Israel showed that level 1 was reached by 70%, level 2 by 50%, level 3 by 30 to 40% and level 4 by only 10 to 30% of the sample. The frequently low performance of respondents in achievement studies suggests that geography education is not successful in achieving its objective. For instance, an Israeli study succinctly summarizes the situation as: "[...] most of the Israeli students enter junior high school with a partial and

fragmented conception of the water cycle and graduate from it with almost the same misunderstandings. This finding is surprising since water is a central issue in Israel's science curricula at the elementary school and junior high levels" (Ben-Zvi Assaraf & Orion, 2005b, p. 371).

Several options to improve the low systems thinking skills of students have been proposed. These especially include specific ways of teaching and learning (e.g. Ben-Zvi Assaraf & Orion, 2010b; Sell, et al., 2006) as well as working with specific media (e.g. Bodzin & Anastasio, 2006; Bundeszentrale für politische Bildung, 2008; Frischknecht-Tobler, Nagel, et al., 2008; Hildebrandt, 2006; Kali, 2003). Results of studies evaluating these proposed measures are difficult to compare due to great diversity in how systemic thinking is defined, operationalized and measured. Additionally, ideas could also be garnered from other areas such as deep learning, which aims at helping students inter alia to "[...] focus[] on relationships between various aspects of the content [...]" (Wilson Smith & Colby, 2007, p. 206) or general teaching/learning research.

Studies in general, however, show mixed results with regard to the effects of an treatment (cp. e.g. Ben-Zvi Assaraf & Orion, 2010b; Bollmann-Zuberbühler, 2008; Falk & Nöthen, 2005; Hildebrandt, 2006; Kaminske, 1996; Orion & Basis, 2008; Ossimitz, 1996; Penner, 2000; Raia, 2005; Rieß & Mischo, 2008b; Thompson & Reimann, 2007). For instance, in a study with 57 eleventh graders in Germany (Kaminske, 1996) students were asked to draw cause and effect diagrams about a geo-system after a corresponding class. Many students merely reproduced facts they learned by heart as well as often gave responses with gaps/mistakes regarding the causality and proportionality of the factors and relationships in the system. Similarly, a study conducted with 102 eleventh to 13th graders in Germany (Hildebrandt, 2006), divided into five groups based on the perspectives (1 or 2) contained in the learning materials and whether they were either not required to work with diagrams, only were the recipient of diagrams or had to construct diagrams showed no significant differences at the $p=0.05$ level in the learning gain from pre- to posttests between groups. A study conducted with 44 eighth graders in Switzerland showed that students participating in a unit training systemic thinking skills demonstrated significantly more system forms of representation in the posttest and a significantly

higher complexity of prediction of effects than the control group (Bollmann-Zuberbühler, 2008). The mixed results with regard to the effectiveness of treatments for the improvement of systemic thinking show that geography education still needs more research to identify the best ways to improve these competencies under various conditions.

2.2 GIS in geography education

Geo-media in the classroom have different functions. These include conveying information about non-accessible areas, helping in the visualization and analysis of geographical topics, facilitating student learning, providing opportunities for the individualized support of students, enabling the acquisition and exercise of media/methods competencies, aiding communication, influencing the students' emotions and affections as well as motivating them and preparing students for the job market and participation in society (e.g. Cremer, et al., 2004; Haubrich, 2006; Klein, 2007; Köck & Stonjek, 2005; Rinschede, 2003; Schallhorn, 2007; Siegmund, 2002). Geo-media are not a substitute for real world experiences, however, since "children need *actual* experiences in order to apply *virtual* technologies with a critical perspective and ethical practice" (Alibrandi, 2003, p. 7).

The terms media, methods, sources of information, working techniques, teaching/learning materials and media carriers partly overlap and are difficult to differentiate, especially when used in context of the competencies necessary to deal with them. Various lists and classifications of media relevant to geography education and media competencies exist (e.g. Bauer, et al., 2007; DGfG, 2002, 2007a; Haubrich, 2001, 2006; Kestler, 2002; Klein, 2007; Köck & Stonjek, 2005; Maier, 1998; Rinschede, 2003; Schallhorn, 2007). Klein (2007, p. 9, translated) defined geo-media as

> "[...] mono- or multi-medial forms of representation for portraying discrete or continuous spatial phenomena and their temporal change. They can – in various degrees of complexity – serve the acquisition, management, analysis and presentation of geo-factors or geo-objects and their geo-data in the integrative interrelationship of physical, biotic and human issues"

and thus could play a crucial role in acquiring geographic system competency.

2.2.1 Extent of GIS use in education

The frequency of use varies considerably between different geo-media. In a study in Schleswig-Holstein (Klein, 2007), secondary school students named atlas/maps as the most frequently used media, followed by geography textbooks. GIS was the least used and most unknown medium (Fig. 5, question: "How often does your teacher employ these media in geography lessons?", translated). From the teachers' perspective, GIS and WebGIS are the least used media, too (Fig. 6, question: "How often do you employ the following media in geography lessons?", translated). There are no significant differences between frequency of use by sex or age of the teacher or by different age levels (grades 5-6, 7-10, 11-13).
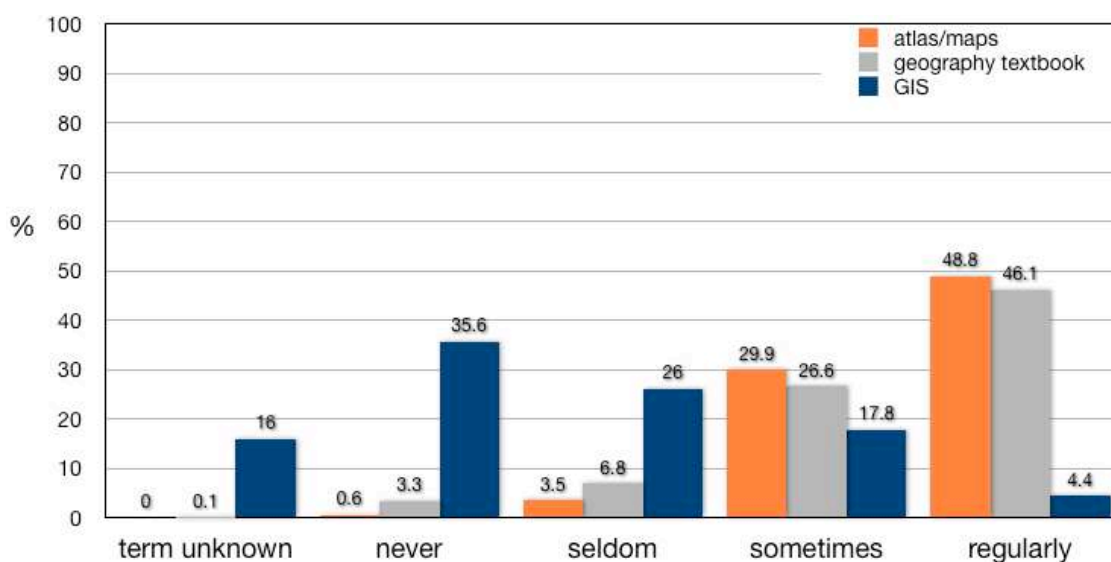
Fig. 5: Frequency of use of geo-media: student answers (%)
    (selection, *n* = 720) (based on Klein, 2007, p. 149)
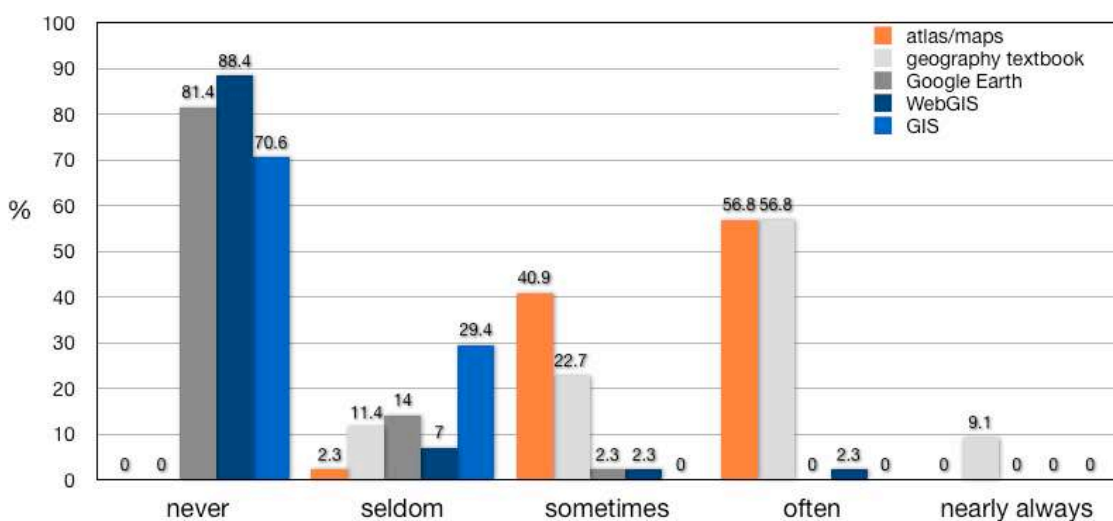
Fig. 6: Frequency of use of geo-media: teacher answers (%)
    (selection, *n* = 44) (based on Klein, 2007, p. 154)

In a study by Höhnle, Schubert and Uphues (2010), 44.7% had never used web-based GIS and 71% had never used DesktopGIS in the classroom, compared to 20.1% for digital globes (n=410, p. 151). Thereby "[y]oung teachers do not use GI(S) more often than older teachers" (ibid, p. 152) and there is no significant difference between teachers who had GI(S) as part of their teacher training and those that did not (ibid p. 154). There are, however, significant effects of self-assessed computer competence, private GI(S) usage, teaching computer science and professional development (ibid, p. 152-155). Moreover, there is a significant correlation to "[...] reading of didactical journals in geography [...]" (ibid. p. 155). There are mixed results with regard "[...] to the existence of other GI(S) using teachers in the teaching staff" (ibid. p. 156) and the teachers' assessment of GI(S)' potential (ibid. p. 156-157).

GIS implementation in German schools started about 1996/97, accompanied by a rise in publications in German geography education journals and developments of GIS software specifically designed for use in schools (e.g. Cremer, et al., 2004; Herzig, 2007; Höhnle, et al., 2010; Schleicher, 2004a). Apart from these special developments, professional programs like ArcView or Idrisi have been used in schools (e.g. Cremer, et al., 2004; Herzig, 2007; Schleicher, 2004a), and free DesktopGIS software such as the GDV Spatial Commander has been advocated, also. In 2003, the first German web-based GIS for schools (www.webgis-schule.de) was made available (Tschirner, 2009), followed by others such as http://webgis.bildung-rp.de, www.sn.schule.de/~gis or http://diercke.webgis-server.de. These options differ considerably with regard to their functionalities and topics, and hence, how they can be used in the classroom. In practical geography education journals a number of class project experiences and lesson ideas have been reported (e.g. Benedikt & Danhofer, 2005; Falk & Nöthen, 2004; Krause, 2004; Püschel, 2004, 2005; Püschel & Schäfer, 2004; Sander, 2002; Unterthurner, 2004; Zürl, 2005).

Internationally, GIS seems to be far from ubiquitously, but increasingly used. For instance, GIS had been used in less than two percent of US high schools (Kerski, 2003). A survey mailed to 1520 high school teachers in the USA already owning a license for either ArcView, Idrisi or MapInfo showed that nearly half of the responding teachers were not using it (Kerski, 2003). In a more re-

cent US study with 70 sixth graders, while over 70% had experience with virtual globes, very few used or even had heard of GIS (Clagett, 2009). In a survey study conducted in the UK, 11.5% of the 243 participating schools were using GIS at the time, 0.8% had given up on GIS and the rest were non-users (Ordnance Survey, 2004). Similarly, in a survey conducted in Finland 11.59% of 69 upper secondary school teachers responding "[...] had already used GIS in the classroom" (Johansson, 2003, p. 3). Moreover, GIS use in school also seems to spread to developing countries. For instance, within a program, 40 secondary schools in Rwanda have been teaching with GIS in 2009 (ESRI, 2010). An overview of the implementation situation in different countries can be found in Milson, Demirci and Kerski (2012). On the whole, however, statistics are far from comprehensive, making it difficult to judge the true extent of GIS use from primary to upper secondary education.

2.2.2 Rationales for GIS use in education

Working with GIS is said to have numerous positive effects on students. However, few of them have been validated empirically, especially in the German context. For instance, GIS use is stated to be motivating and interesting (e.g. de Lange, 2006a; Falk, 2004; Heiken & Peyke, 2005; Lund & Sinton, 2007; Porsch, 2005). It is also said to foster/enable a wide variety of skills and competencies, such as methods competency, general computer and media literacies, spatial thinking skills and spatial awareness, map competency, personal responsibility, self-directed learning and social competency, critical thinking skills, (artistic) creativity and reading literacy (e.g. Cremer, et al., 2004; de Lange, 2006a; Falk, 2003, 2004; Falk & Hoppe, 2004; Falk & Schleicher, 2005; Feyk, 2006; Heyden, 2004; Joachim, 2006; Lund & Sinton, 2007; Porsch, 2005; Schäfer, 2002; Schäfer, 2005; Schleicher, 2006; Schwab & Kussmaul, 2005; Sinton & Bednarz, 2007; Thyne, 2005; Zürl, 2005). It is claimed that in general, GIS use would improve the learning process e.g. making it more lasting and effective (e.g. Falk & Hoppe, 2004; Falk & Schleicher, 2005), as well as simplify and augment geographic investigations (e.g. Heiken & Peyke, 2005; Malone, Palmer, & Voigt, 2002; Sinton & Lund, 2007). GIS is also stated to change the way content is taught and learned, for instance by making teaching more col-

laborative and inquiry-based (e.g. Alibrandi, 2003; Johansson, 2003) or ena-
bling a different way of analysis e.g. also for local examples (Koller, 2005;
Schwab & Kussmaul, 2005). Working with GIS is said to improve the under-
standing of complex contents, interactions and processes, both in the envi-
ronment and the society, as well as systemic thinking and subject competen-
cies (e.g. Bodzin & Anastasio, 2006; Cremer, et al., 2004; de Lange, 2006a;
Falk, 2003, 2004; Falk & Hoppe, 2004; Falk & Nöthen, 2005; Falk & Schleicher,
2005; Heiken & Peyke, 2005; Herodot, 2009; Lund & Sinton, 2007; Schäfer,
2002; Sinton & Bednarz, 2007). Moreover, it is argued that "[t]he vast amounts
of information available today require powerful tools like GIS to help people de-
termine what it all means" (Malone, et al., 2002, p. xxiv). De Lange (2006b, p.
13, translated) even claims that the competencies specified in the common re-
quirements for the *Abitur,* the high stream school leaving examination, "[...] can
be reached optimally through the use of GIS". With GIS use increasing globally
in a wide variety of fields and a growing demand for trained GIS specialists (cp.
e.g. Alibrandi, 2003; ESRI, n.d.; Fitzpatrick & Dailey, 2000; Gewin, 2004; Glover,
2006), GIS is also claimed to prepare students for the job world (e.g. Cremer, et
al., 2004; Falk & Nöthen, 2005; Falk & Schleicher, 2005; Joachim, 2006; Zink &
Scheffer, 2009). Moreover, Falk (2004, p. 192, translated) argues that "[t]he in
the university sector by now natural use of geographic information systems
and their great importance beyond the subject boundaries should be also ac-
commodated by the school geography education".

Beyond the potential positive aspects on students, another line of argument is
that students can expect GIS to be used in school education due to its preva-
lence in every day life (e.g. de Lange, 2006a; Falk & Hoppe, 2004; Joachim,
2006; Reitz, 2005; Siegmund & Naumann, 2009; Zürl, 2005). GIS technology
based examples include navigation systems, route planners, 'find a branch'-
services, the TV weather report and interactive city maps (e.g. Alibrandi, 2003;
Naumann, et al., 2008; Schleicher, 2004a; Siegmund & Naumann, 2009). Con-
sequently, it has been claimed that "GIS is used daily in so many aspects of
human activity that it will become one of those skill sets as basic as word
processing is today. Looking historically at the uptake of computer use in the
classroom, this transition is roughly comparable to the status of word process-

ing integration in 1990" (Alibrandi, 2003, p. xi). Thus GIS is seen to prepare students for the world they live in (e.g. Alibrandi, 2003; Herodot, 2009).

2.2.3 Potentials of GIS use for improving student achievement

Which features of GIS use lead to a positive effect on students' achievement is not yet understood. Possible features include interactivity (e.g. Hall, 2000; Johansson & Pellikka, 2005; Schwarz, 2005; Tillmann, 2006), the layer principle and the possibilities for adapting complexity (e.g. Falk & Nöthen, 2005; Krause, 2005; Pönitz, 2008) or capabilities for analysis (e.g. Cremer, et al., 2004; Falk, 2003; Falk & Nöthen, 2005; Kerski, 2000). For instance ESRI (1998, p. 2) explain the functionality of a GIS using an overhead projector with transparencies as example and go on to state that

> "Standing before the overhead, you mix and match the layers at will, magically changing classification schemes and modifying symbols, colors, patterns, and combinations. You can zoom in and out, seeing all the information available or only the data you specify, comparing this layer with that feature, exploring the data in every way imaginable. As you play with these layers of information, relationships appear."

The last sentence might point especially to the potential of GIS use for improving systemic thinking. However, as the crucial features have not been empirically identified yet, it is unclear whether simple web-based GIS specifically created for education or professional DesktopGIS are needed, and which differences in impact and use exist (see also e.g. Baker, et al., 2012, p. 274-275).

Worldwide, the number of empiric studies dealing with the impact of GIS use on student achievement is growing and already spans a wide variety of contexts and objectives (cp. e.g. Baker, et al., 2012; Huynh, 2009). The number of empiric studies is especially limited with regard to Germany. Available studies conducted in Germany and worldwide have employed a wide variety of methodologies and measurement instruments, which makes it somewhat difficult to compare results across studies. Because of the complexity of trying to determine the impact one medium has on achievement and the hitherto rather mixed study results, more research is needed (e.g. Baker, et al., 2012; Bednarz, et al., 2013), as well as a drawing together of different studies into a coherent theory.

With regard to non-quantitative studies, papers such as Shin (2006), Keiper (1999), Doering & Veletsianos (2008), Milson and Earle (2007) or Aladag and Aladag (2008) show that GIS can help both primary and secondary students to improve aspects of their geographic competencies. Other studies (e.g. Drennon, 2005; Favier & Van der Schee, 2009; Wigglesworth, 2003) seem to indicate a strong influence of prior experience or background knowledge. For instance, a study with high stream students in Germany (Falk & Nöthen, 2005, p. 51, translated) shows that while GIS helped them to look at a topic from multiple perspectives and gain some insight into interdependencies, they "[o]ften [...] lacked important background knowledge to comprehend "the structure, the function and the interaction in the individual systems"" and thus the project "[...] could only moderately contribute to fostering systemic thinking" (see also Falk, 2004).

Studies having one test data point also present a diverse picture. Some studies using a one group design show that the majority of students agree that GIS helps them with regard to visualizing or understanding geographic concepts (e.g. Klein, 2005; Storie, 2000). Yet, in another study by Klein (2007), the percentage of students who reported that they understood nothing at all when using GIS (6.8%) was the highest of all media included in the survey. An additional 10.2% reported bad understanding when using GIS. Flecke (2001) conducted a small study in Germany after a short GeoMedia Professional unit with a quiz containing both technical and content questions and some evaluation questions shows mixed results. However, the study has large methodological problems (e.g. no n reported).

Several studies using a two group design noted that the GIS group achieves higher scores than the control group, which often consists of students using conventional methods such as paper maps (e.g. Aarons, 2003; Baker, 2002; Baker & White, 2003; Demirci, 2008; Patterson, Reeve, & Page, 2003). In contrast, some studies also show no significant difference (e.g. Clagett, 2009) or a higher achievement of the non-GIS group (e.g. Meyer, Butterick, Olkin, & Zack, 1999).

Similar mixed results can be observed in pre-/posttest studies. Sometimes one group designs seem to show an improvement from the pre- to the posttest (e.g.

with regard to content knowledge, Fun, 2005), while at other times, no significant improvement was found (e.g. with regard to self-reported understanding of the work they do in the subject, West, 2003). With regard to two (or more) group designs, studies such as Lee (2005), Aladag (2007) or Songer (2010) showed that GIS had a positive impact on posttest scores compared to a non-GIS group. Moreover, studies such as Wanner and Kerski (1999), Kerski (2000) or Simmons, Wu, Knight and Lopez (2008) show that GIS helped students as measured by the difference between pre-and posttest compared to a control group with regard to some skills tested but not with regard to others. Additionally, studies such as Crews (2008) or Hagevik (2003) seem to indicate that the impact of GIS use on student achievement might differ according to how and how much the use is implemented.

An overview of research in GIS education can be found e.g. in Baker et al. (2012).

2.2.4 Costs of GIS use in education

The implementation of GIS into education is not without costs. Costs come for instance in the form of time, money and 'hassle' (cp. e.g. Aladag, 2009; Audet & Paris, 1997; Johansson, 2003; Kerski, 2000, 2003; National Research Council, 2006; Schleicher, 2004b; Wiegand, 2006; Zink & Scheffer, 2009). Time costs include training time for the teacher (GIS specific and general computer skills), training time for the students, preparation for the lessons/creating materials, development of a GIS didactics (e.g. how to apply GIS functionality to teach geographic concepts) and a portion of the limited class time. Besides the time costs, direct monetary costs include training courses, materials, software, hardware, and in some cases, data. 'Hassle' costs refer to the added effort that is often still needed to include GIS into instruction. Examples are booking the computer room in competition with other subjects' teachers, swapping lessons with another teacher to have more than 45 minutes to work on a project at one time or convincing the computer lab administrator to install software and provide technical support.

Costs differ between various GIS programs and between various ways of implementation, such as Web-based vs. Desktop-based and prefabricated material vs. creating one's own materials. In general, working with prefabricated student worksheets that accompany Web-based GIS services designed for

schools or that accompany, together with a ready-to-use-data package, a DesktopGIS is associated with less costs than creating one's own materials. For instance, for a DesktopGIS application, the teacher might have to – in addition to creating the student worksheets and after installing the program and getting it to run within the school's computer lab – locate data on the web, translate and simplify the attribute information and change the projection in order to make it fit to other layers in the data set. Consequently, the new generation of geography text books, if including GIS, normally does so in the form of simple GIS-based web applications (e.g. Engelmann, et al., 2005; Flath & Kulke, 2007; Obermann, 2005). However, even these simple applications are not free of costs, as, for instance, non-intuitive user interfaces and functionalities require time to learn.

What matters, however, are not only the absolute costs, but much more, the cost-effectiveness. Cost-effectiveness analyses could give guidelines as to which forms of implementation provide the highest impact on the achievement of desired outcomes relative to their different forms of costs for all involved entities (Levin, 1995, 2001; Loi & Ronsivalle, 2005), comparing different forms of GIS implementation and other newly propagated as well as traditional means. To date, such explicit, comprehensive and empirically supported cost-effectiveness or cost-benefit analyses seem to be very rare.

2.3 Role of student variables

In some situations, average scores can be useful. Due to the complexity of educational processes, however, a more differentiated view is often needed. For instance, averages do not take into account whether certain students are especially benefited or disadvantaged by an treatment. Achievement gaps and factors influencing the probability for high achievement have been studied both on the level of the individual student, the class, the teacher, the school, an area and the school system as a whole on different temporal scales (cp. e.g. Ferguson, 2002; Gandhi Kingdon, 2006; Hattie, 2003, 2009; Kfir, 1988; Ladson-Billings, 2006; NSSE, 2006; OECD, 2006a). Thereby, a multitude of variables being potentially influential have been identified. In general, due to test time considerations, only a limited number of them can be included in any one study.

2.3.1 Age

It could be assumed that older students have a higher achievement in geographic system competency than younger students. Reasons for this are, for instance, the generally increasing maturity, background knowledge, and experience. However, this seems to be the case only to a limited extent. Studies such as Sweeney & Sterman (2000) or Kainz & Ossimitz (2002), which were conducted with adult learners, did not show any systematic variation of systems thinking skills by age. Similarly, also studies including both teenagers and adults, such as that of Ossimitz (1996) or Kasperidus et al. (2006, particpants aged 15 to over 30) showed no significant relationship between age and systemic thinking. Also Ben-Zvi Assaraf and Orion (2010b), state that "[c]omparing the initial abilities of these students (8th grade students) with the pre-test outcomes of the current study (4th grade students) points to a quite similar level of abilities" (p. 558). In contrast, the study of Sommer (2005), conducted with primary school children, showed that older students had a higher achievement in systemic thinking skills than younger students. An interesting finding is provided by Hagevik (2003): she found that older students scored significantly lower than younger students on an environmental content test. This might be explained by retained students.

For similar reasons, it could also be assumed that the impact of GIS use on student achievement differs by age. Huynh (2009), for instance, in a study with secondary school and university students identified three different groups: novices (aged 14 to 22), intermediates (aged 18 to 24) and experts (aged 22 and above). These differed not only in their geospatial score and geographic skills, but also in their GIS problem-solving score: out of 15 possible points, experts had a mean of 12.3, intermediates 10.5 and novices 8.3. Moreover, the strategy employed by the three groups differed. While novices employed a visual, trial and error strategy, intermediates used a combination of visual and deduction strategies and experts a structured, logical deduction strategy (cp. also a study with teachers, high school students and GIS experts, Audet & Abegg, 1996; a study with sixth and seventh graders, Wigglesworth, 2003). In a study with seventh and eighth graders, older students scored significantly higher with re-

gard to GIS analysis than younger ones (Hagevik, 2003). Nevertheless, it also needs to be taken into account that GIS has been implemented across a wide age range (cp. e.g. Bodzin, 2008, for a study with primary school students; Wilder, Brinkerhoff, & Higgins, 2003, for a study with teachers).

2.3.2 Sex

Differences between male and female learners have been studied in different areas, and are often placed in a nature-nurture-framework of explanation. It appears that males, in many countries, generally have a higher achievement in geography and earth science (e.g. Bednarz, et al., 2013; Foshay, Thorndike, Hotyat, Pidgeo, & Walker, 1962; LeVasseur, 1999; Martin, et al., 2008; National Center for Education Statistics, 2002, 2011; OECD, 2007c; Winship, 2004) and systemic thinking (e.g. Kainz & Ossimitz, 2002; Kasperidus, et al., 2006; Ossimitz, 2001; Sweeney & Sterman, 2000). For instance, the PISA 2006 scale "earth and space systems" shows that, in general, "[...] males tend to outperform females [...]" (OECD, 2009, p. 37), as well as having "[...] a greater variation of performance than females, that is, they tend to have comparatively higher proportions of top performers but also of students at risk" (OECD, 2009, p. 37). There are, however, also some studies that show no significant difference in most tasks or even a lower achievement of males (e.g. Bednarz, et al., 2013; Butt & Weeden, 2004; Foshay, et al., 1962; Huynh, 2009; LeVasseur, 1999; Maierhofer, 2001; Martin, et al., 2008; Montelllo, Lovelace, Golledge, & Self, 1999; OECD, 2007c; Ossimitz, 1996; Sommer, 2005). For a discussion of some sex difference aspects in the German situation see e.g. Hemmer (1995). Moreover, sex differences with regard to spatial abilities within and beyond the geo-sciences have been frequently documented (e.g. Baenninger & Newcombe, 1989; Barke & Engida, 2001; Montelllo, et al., 1999; Quaiser-Pohl, Geiser, & Lehmann, 2006; Self & Golledge, 1994).

Regarding GIS use, existing studies seem to indicate that the improvement in achievement or the posttest scores after a GIS treatment respectively mostly do not differ between male and female learners (e.g. Clagett, 2009; Hagevik, 2003; Kerski, 2000; Kinzel, 2009; Lee, 2005). However, there are also studies

that show a higher overall achievement of females in a GIS course (e.g. Clark, Monk, & Yool, 2007).

2.3.3 Pre-achievement

In general, pre-achievement has a large influence on achievement (e.g. Hattie, 2003; Shapiro, 2004). Traditionally, in the German school system, students have been usually streamed into different school types according to their achievement in grade four. Sometimes, new legislation, however, allows for the parents to choose freely (e.g. Wetzel, 2014). This initial difference in achievement as well as what Malcom Gladwell (2008, p. 25) explains as

> "[i]f you make a decision about who is good and who is not good at an early age; if you separate the 'talented' from the 'untalented', and if you provide the 'talented' with a superior experience, then you're going to end up giving a huge advantage to that small group [...]"

lead to students in a higher school type generally outperforming those attending a lower type (e.g. Hüttermann, 2004; Prenzel, et al., 2008; Uphues, 2007). For instance, in a study with 424 sixth graders dealing with systemic thinking in a biology context, high stream students had a significantly higher mean than mid stream students and there was a significant correlation between biology grade and systemic thinking achievement (Rieß & Mischo, 2008a).

School type differences have been documented in a range of areas of geo-science education, such as risk perception (Fiene, 2014), attitudes to globalization (Uphues, 2007), interest (e.g. Hemmer & Hemmer, 1997) and map competency (Hüttermann, 2004). With regard to digital geo-media, Ditter (2013) found for instance that high stream students significantly improve in self-efficacy pre- to posttest after a unit using remote sensing (d>0.2), but mid stream students do not (d<0.2). In his study, stream also has a significant relationship to motivation/self-efficacy/interest cluster membership (crosstabs). In a study of students' satellite image use, Siegmund (2010) also included mid stream vs. high stream membership as one of several variables in the cluster formation. Uphues (2007) also found that in two items, more high stream than low stream students took into account both the effects on Germany and on the other country.

Pre-knowledge not only provides something one can use when solving a task, it provides also anchor points for new information (cp. e.g. Smith, Gordon, Colby, & Wang, 2005; Sommer, 2005; Taber, 2001). Thus, pre-knowledge and pre-experience have an impact on learning and achievement (e.g. Hammann, Phan, Ehmer, & Grimm, 2008; Phan, 2007; Schmitz, 2006; discussion in Sell, et al., 2006). For instance, Sommer (2005) showed that biological pre-knowledge had a significant influence on achievement of aspects of system competency. Additionally, studies such as Kali et al. (2003) or Sibley et al. (2007) show that achievement in systemic thinking can also be influenced by the context used. Maierhofer (2001), in a biology study dealing with systemic thinking, also found that while students' scoring low on the pretest (both with regard to number of system variables used and quality of relationships) improved significantly, high scoring students did not. Hildebrandt (2006) found differences between low and high pre-knowledge students in terms of which way of working with diagrams was beneficial for improving understanding.

With regard to the impact of GIS use on achievement there seem to be no studies yet that compare different German school streams. While on the one hand, GIS is primarily included for the high stream as obligatory into school curricula in Germany (e.g. Siegmund & Naumann, 2009), on the other hand an interesting finding has been presented in a study conducted in the USA by Kerski (2000). He found that while "A" students improved both with GIS and with traditional materials, "C" and "D" students only showed significant gains with GIS.

A study with secondary school students in Germany showed that there were significant differences between students belonging to different types of PC users and the subjective learning success with computers and computer-based media such as GIS, with students classified as "PC inexperienced" having the lowest value (Klein, 2007).

2.3.4 Language and migration background

Regarding the achievements of students of different ethnic and linguistic backgrounds in a particular school system, several studies have shown (some) minority student groups to be at a disadvantage (e.g. Brind, Harper, & Moore, 2008; Fashola, Slavin, Calderón, & Durán, 1997; Ladson-Billings, 2006; Nu-

sche, 2008; OECD, 2006b; Zuzovsky, 2007), including in geography education (see e.g. LeVasseur, 1999). This problem is especially pronounced in Germany. For instance, the PISA 2006 science competency results (which included some earth science items) showed a large gap in achievement between children with and without migration background (OECD, 2007a). In general, PISA showed that Germany had one of the largest gaps of all participating countries between children with migration background and those without. The study also showed that second generation children on average score lower than young people that have migrated within their own lifetime (e.g. OECD, 2003b; Prenzel, et al., 2006). While digital reading performance is not reported for Germany, "[a]cross OECD countries, the pattern of results indicates that native students perform at a higher level than their immigrant counterparts" (p. 127) and similarly, students "[...] whose language at home is different from the assessment language [...]" score less than "[...] students whose language is the same as the assessment language" (p. 127) (OECD, 2011).

Language is crucial for thinking and formal education. Ontological research, for instance, showed that different languages use different categories for geographical features (e.g. Mark & Turk, 2003). On the one hand, the achievement gap in some studies can be explained by testing content in the majority language that the children do not yet know well enough (e.g. Escamilla, Chávez, & Vigil, 2005). On the other hand, in Germany, even adolescents with migration background who predominantly use German in their every day life did not reach the competency level of children without migration background (e.g. Prenzel & Deutsches Pisakonsortium, 2005).

Contrary to these findings, some research seems to indicate that bilingual children have a cognitive advantage over monolingual children, including in abilities such as thinking abstractly about language, classifying, reasoning by analogy, having mental flexibility, or thinking divergently (e.g. Lee, 1996). With regard to an advantage in subject areas, research results have been mixed (e.g. Taylor-Ward, 2003). Some studies have shown bilingual children who learn both their languages for instance in dual language or two-way programs to have greater achievements than monolingual students or students less proficient in their first language in areas such as science, writing, mathematics, reading, ana-

lytical reasoning or spatial abilities (cp. e.g. Escamilla, et al., 2005; Laib, 2007; Lee, 1996). Furthermore, while some research indicates that which languages the child speaks is not important, but only the level of proficiency, other research seems to be pointing to an impact of whether the child's first language is valued or not (e.g. Bialystok, 2008; Laib, 2007; Lee, 1996; Marlow, 2008).

In contrast to the substantial general educational literature on the influence of linguistic, ethnic and migration background on student achievement, it seems to have hardly been included in studies dealing with the impact of GIS use.

2.3.5 Affective variables

The affective domain is about a variety of constructs, including "[...] student attitudes, interests, appreciations, values, and emotional sets" (Mililani Mauka Elementary School, n.d.), self-esteem (NCREL, n.d.), "[...] beliefs, opinions [...] and motivation" (Koballa, n.d.).

Student characteristics, such as beliefs, attitudes, self-efficacy, self-concept, motivation and interest play a crucial role in student achievement (e.g. Duit & Treagust, 2003; e.g. Edelmann, 2000; Heinze, Reiss, & Rudolph, 2005; OECD, 2007b; Pastorelli, et al., 2001; Rider & Colmar, 2005; Roberts II, 2003; Uwah, McMahon, & Furlow, 2008; Vogt, 2007; Woolfolk Hoy, 2004). For instance, Sommer (2005) showed a significant influence of interest in the subject on achievement in a system competency pretest and a significant influence of interest in the topic on achievement in a test during the unit. Studies have reported mixed results in sex differences with regard to self-concept and self-efficacy (e.g. Baker & White, 2003; Neria, Amit, Abu-Naja, & Abo-Ras, 2008; Pastorelli, et al., 2001; Preckel, Goetz, Pekrun, & Kleine, 2008). Furthermore, mixed results in self-efficacy and self-concept have also been found by country of origin or ethnicity (e.g. Lee, 2009; Neria, et al., 2008; Pastorelli, et al., 2001).

Within geography education, differences in students' interest in different topics, geographical areas and media/methods as well as differences by sex have been observed to some extent (e.g. Bayrhuber, et al., 2002; Hemmer, et al., 2007; Hemmer & Hemmer, 1997; Hemmer & Hemmer, 2002a; Hemmer, 2000, 2008b; Klein, 2007; Obermaier, 1997, 2002a, b; Reinfried, 2006; Schleicher, 2002a). For instance, Klein (2007) found that students had a mean interest of

3.07 in GIS on a scale ranging from one (no interest) to five (high interest), with boys having significantly higher interest than girls. With regard to atlas and maps, mean interest was 3.44, again with boys scoring higher. In contrast, movies had a mean score of 4.32 with no significant gender differences. Additionally, some research seems to indicate a difference in geographical interest by school type (e.g. Hemmer & Hemmer, 1997; Hemmer, 2008b).

The results regarding the impact of GIS use on these kind of student characteristics seems mixed. In West's (2003) study scores on the "attitude to subject" scale declined after using GIS, while the scores on the computer related scales (e.g. perceived control of computers) increased. Baker and White (2003) found that the GIS group showed a significant increase in their attitudes toward technology and science self-efficacy while students in the traditional mapping group significantly increased their attitudes toward science. Crews (2008) showed that there were no overall significant differences "[...] in student science interest scores pre-post between treatment and control groups [...]" (p. 103). However, he also noted instructor effects with regard to the students' interest in science and technology, with overall scores decreasing for the teacher with a high implementation of geospatial technologies and scores increasing for the low implementation teacher.

2.3.6 Classroom setting

Research has shown that besides the individual, the teacher, the class as well as peer group effects have an impact on individual achievement (e.g. Betts, 2004; Burke & Sass, 2008; Hattie, 2003, 2009; Heinze, et al., 2005; Sacerdote, 2001; Weinert, 2001). For instance, in the study conducted by Baker and White (2003) instructor effects explained 20% of the variance in achievement while GIS or control group membership only explained 2%.

Moreover, on a higher level, although an overarching framework exists, Germany's federal states have considerable freedom in organizing their educational system and curricula. Consequently, comparative studies show large differences in student achievement between the individual federal states, not just overall but also with regard to specific sub-groups such as children with migration background (e.g. Bos, et al., 2008; Helmke & Hosenfeld, n.d.; MPG, n.d.;

Prenzel, et al., 2004). Thus, results from one federal state might not be easily generalizable to others.

2.4 GIS use and geographic system competency

While there is a considerable amount of theoretical deliberations regarding geographic system competency, only a limited number of empirical studies have been conducted, especially in Germany. Consequently, much research is still needed regarding the achievement of the guiding objective of geography education in Germany as well as regarding the factors influencing achievement and the effectiveness of treatments to increase achievement (cp. e.g. Ben-Zvi Assaraf & Orion, 2005a, p. 557 for the international situation). The lack of empirically validated competency models impedes the exploration of the influences of GIS use on competency development.

While GIS has been claimed to have numerous positive impacts on the students' working with them, including fostering their systemic thinking skills, this is not yet supported by a sound research base. GIS education is of increasing interest to researchers worldwide, and consequently, a growing amount of evidence regarding the effectiveness of GIS use for improving student achievement exists (cp. e.g. overview in Baker, et al., 2012). Yet, the existing research map has still many white spots. Additionally, research findings from other countries, while providing valuable insights, need to be evaluated with regards to their applicability in other national contexts (see also e.g. Aladag, 2009). Consequently, there is a need to empirically study the impact of GIS on aspects of one of the central aims of geographic education. Especially needed are studies employing advanced methods such as IRT or multi-level analysis, which are still underrepresented in the German geography education research literature in general, and German GIS education research literature in particular, as the review of studies above showed.

3 Charting the course

To answer the questions at the core of this dissertation comprehensively, one would need to examine differences between learning with GIS and paper maps for a multitude of topics, types of GIS/ paper maps and methods with a great diversity of students and teachers. This would need to be done in a way which is both adequate to the situation and producing results that are comparable across these different settings. This is beyond the scope of this dissertation due to resource constraints, which also influence, for instance, sample size, methodology and treatment topic.

Similar to other studies (see ch. 2), this dissertation uses a pre-/posttest design with comparison group. A comparison group is often used in studies seeking to explore effects of a particular treatment (cp. e.g. Verma & Mallick, 1999) and has advantages with regard to the estimation of treatment effects compared to single group pre-/posttest studies (cp. e.g. Grimshaw, Campbell, Eccles, & Steen, 2000). Similar to existing studies (e.g. Baker & White, 2003; Clagett, 2009; Kerski, 2000), traditional mapping was used for the comparison group.

In general, research questions can be studied either qualitatively, quantitatively or by a combination of the two (e.g. Bradshaw & Stratford, 2000; Cohen, Manion, & Morrison, 2000; Haslam & McGarty, 2003; Lamnek, 2005; Marshall & Rossman, 1999; Schostak, 2002). The studies conducted for this dissertation are quantitative in nature. The studies use written questionnaires. General advice on how to construct tests/ questionnaires can be found for instance in Moosbrugger & Kelava (2007), Rost (2004), Rost (2005a), Bühner (2004), Cohen, Manion & Morrison (2000) or Haslam & McGarty (2003) as well as in the overviews in other studies such as Uphues (2007).

Due to the necessity of first having to develop an instrument (see ch. 5), this dissertation comprises five studies (Fig. 7). The main study (5) was preceded by four smaller studies. Of these, one was a study with comparison group (3). The other three studies (1, 2, 4) did not include a comparison group since their focus was on test instrument development. The studies were intertwined with an iterative process of test and learning unit development, where feedback

from a previous study, such as Study 1 pointing to explicitly spatial (maps) and not explicitly spatial tasks maybe being something different, led to adaptation of the materials used for the next study.
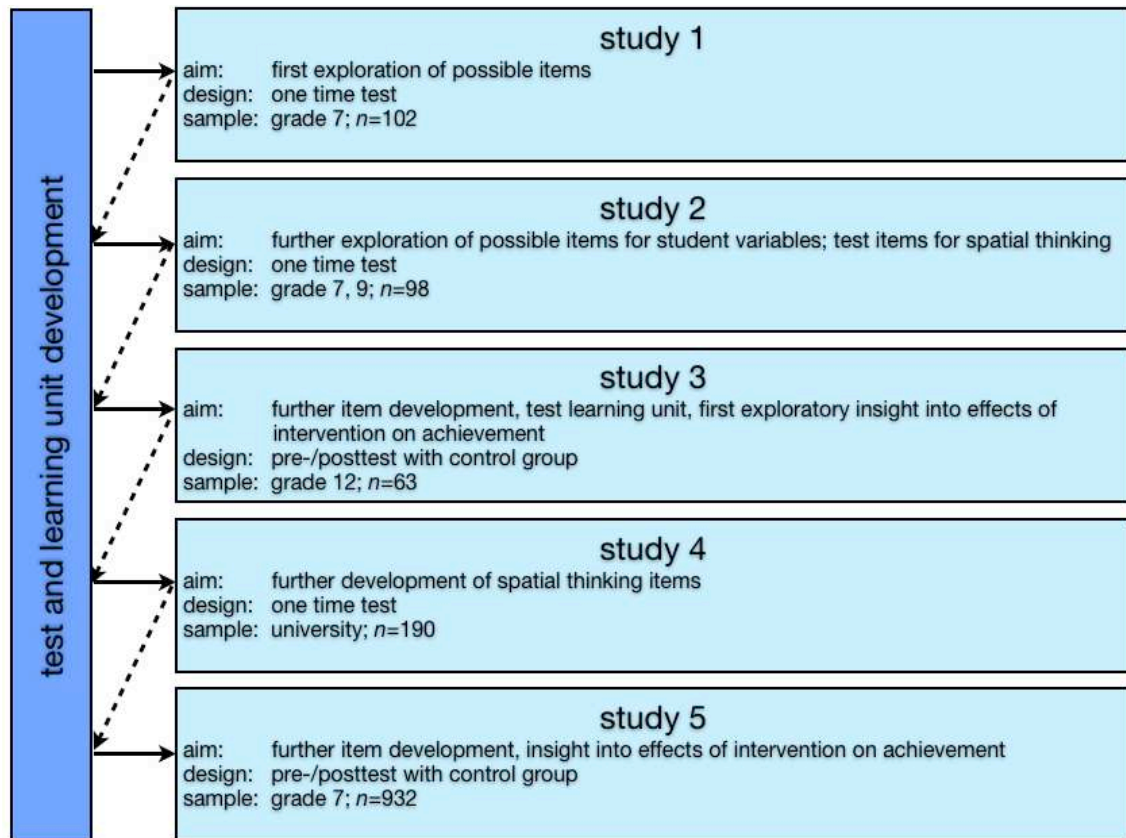
Fig. 7: Study overview including *n* used for analysis

The design of the main study (5) is quasi-experimental in nature, since participants were not randomized, neither in terms of their selection nor in their assignment to either GIS or map group (see e.g. Sapp, 2006). While in study 3, assignment of the students from different classes to either GIS or map was done by one teacher, in study 5 teachers decided whether they wanted to participate with the whole class as either GIS or map group. On the one hand, this constitutes a severe limitation, since it has been argued that "[...] the estimate of effect cannot be attributed to the intervention with confidence due to the non-randomized control group" (Grimshaw, et al., 2000, p. S13). Moreover, it leads to the potential of the sample being biased. Firstly, in general it could be expected that only certain types of teachers are willing to participate voluntarily in a scientific study. Secondly, teachers might have various reasons for selecting participation either in the GIS or the comparison group and thus might dif-

fer from each other (cp. e.g. also Slavin, 2008). However, due to the current policy framework, this study needed to rely solely on the willingness of teachers and students to participate. Consequently, forcing a randomized group assignment on a teacher would have led to some teachers not participating and thus would have also potentially biased the sample. Furthermore, although preferring randomized studies, Slavin (2008, p. 8) also notes that

> "[t]he evidence to date suggests that quasi-experimental studies in which experimental and control groups are well matched, and in which covariates that correlate strongly with pretests (e.g., achievement pretests) are used to adjust outcomes, produce good, if not perfect, estimates of program outcomes, as long as there are no possibilities of selection bias at the individual student level".

To choose one or several topics for the treatment grade level and stream(s) were important considerations. The main target is grade seven, although other students participated (Fig. 7). About GIS use with younger students, less was known in the German context than about older students (grades eleven to twelve/13). This stood in contrast to international research, the national educational standards, curricula of a number of federal states and some of the didactical models for GIS integration, which already include GIS before grade eleven (for overviews see e.g. Baker, et al., 2012; Siegmund & Naumann, 2009; Siegmund, Viehrig, & Volz, 2008; Viehrig & Siegmund, 2012). Curricula play an important role in grade selection. In Baden-Wurttemberg, GIS was first required in the standards for grades seven/eight (for the high stream only, MKBW, 2004a). Moreover, in grade seven differences between school types are likely to become even more apparent than in grade five or six (which are sometimes called "orientation stage"). Additionally, discussions showed that an investigation in grade seven is likely to be easier than higher grades because it is far away enough from the school leaving examinations and similar conflicting demands on classes.

Regarding school streams, GIS use seemed to be most frequent for the high stream both with regard to curricula (MKBB, 2002; MKBL, 2006; MKBR, 2006a, b; MKBW, 2004a, b; MKBY, 2001, 2004; MKHE, n.d.-a, b; MKHH, 2003, 2004; MKMV, 2002a, b; MKNS, 2008a, b; MKNW, 1993, 2004; MKRP, n.d.; MKSA, 1999, 2003; MKSH, 1997; MKSL, 2002, n.d.; MKSN, 2004a, b; MKTH, 1999a,

b; see also e.g. Siegmund, et al., 2008; Viehrig & Siegmund, 2012) and materi-als available (e.g. integration into school books or online lesson plans, such as those found on www.webgis-schule.de). Moreover, GIS use is demanded by the National Educational Standards for the Intermediate School Certificate (DGfG, 2008) and its use was thus likely to be expanded in the middle stream. In contrast, GIS for the low stream seemed to be hardly discussed. Conse-quently, the high and mid streams were chosen for the research.

	study 1	study 2	study 3	study 4	study 5
demographic variables					
sex	x	x	x	x	x
age	x	x	x	x	x
stream/type of geography study	x	x	(only one)	x	x
grade/semester	(only one)	x	(only one)	x	x
languages	x	x			x
migration background		x			x
pre-experience variables					
Kenya	x				(x)
GIS	x	x			(x)
maps					
computer skills	x				
attitude/motivation/interest/ etc. variables					
Kenya	x	x			
GIS and other media	x	x			
geography/geographic systems	x	x	x		
achievement variables					
systemic thinking	x		x		x
spatial thinking		x	x	x	x
treatment variables					
group			(x)		x
feedback			((x))		x
material					(x)

Tab. 1: Variables included in the five studies

The first studies included a greater number of variables, such as interest in Kenya or working with GIS (see Tab. 1). These could not be included in study 5 due to test time considerations, especially after inclusion of spatial thinking tasks. To enhance readability, this report focuses on the development of vari-ables used to test the research hypotheses for study 5. Studies 1 and 5 also featured the time used for the test, however, this is not included as a variable.

3.1 Research questions and hypotheses

The central question of this dissertation is in how far a learning unit using GIS helps students to improve their systemic thinking. The main hypothesis is therefore:

I On average, there will be an improvement in systemic thinking achievement for participants working with GIS in the unit.

Because there are considerable costs associated with the implementation of GIS use (see ch. 2.2.4), it is of interest whether the improvement in systemic thinking is greater or smaller than with a paper map unit. Existing studies show mixed results (see ch. 2). Yet, there are many claims surrounding GIS's potential for competency achievement (e.g. de Lange, 2006b). Within the limitations of this study, therefore, the following hypothesis is tested:

II On average, there will be a greater improvement in systemic thinking achievement for participants working with GIS in the unit than for participants working with paper maps.

There are numerous studies examining the differences between different groups of students (see ch. 2.2.3 and 2.3), and whether or not there are interactions between the treatment and other variables (e.g. does one treatment work better for one group than for another). However, in the area of GIS, studies systematically examining many of the possible effects are still largely lacking and possible influencing factors are often ignored in the "GIS promise literature". Moreover, factors influencing the achievement and improvement of systemic thinking in the geosciences are not yet fully explored. Based on the literature review in ch. 2.3 and the focus of this report as outlined above, the following hypotheses are tested:

III Improvement of achievement:

On average ...

a) ... there is no effect of age on improvement.

b) ... there is no effect of sex on improvement.

c) ... migration background has a negative effect on improvement.

d) ... there is no effect of language background on improvement.

e) ... pre-achievement has an effect on improvement.

Additionally, the results of study 1 hint at there being a difference between tasks that are graphics based and those that are not. This led to an exploration of the literature on spatial thinking and subsequently an inclusion of basic spatial thinking tasks as a possible variable in later studies. Furthermore, after the first feedback in study 5 it was decided to do another version of the materials. Therefore, the following hypotheses are also tested:

III Improvement of achievement:

On average ...

f) ... there is no effect of pre-spatial thinking score on improvement.

g) ... there is no effect of material type on improvement.

Pre-experience with the medium is a potentially important variable and was included within the materials in study 5, however, very few of the students had previously worked with an educational GIS program, while nearly all students specified having already worked with an atlas (see ch. 6). Consequently, the hypothesis

III Improvement of achievement:

On average, ...

h) ... pre-experience with an educational GIS program has a positive effect on improvement.

can only be tested in a very exploratory manner.

Due to the complexity of the educational reality, the setting (e.g. teacher, classroom) cannot be controlled especially if a large number of students are to be examined and thus a near-laboratory situation is not feasible. Additionally, the 'GIS promise'-literature generally seems to refer to normal school instruction, not specifically controlled laboratory conditions.

3.2 Statistical considerations and data analysis

To statistically examine the properties of the items and scales, different guidelines and methodologies exist. Thereby, with regard to some of the issues relevant to the analysis of the present studies, differences in interpretation are common. This also includes the issue of sample size recommendations for the different statistical procedures.

3.2.1 Scale quality and analysis

In general, it is recommended not to use only a single item but scales, for instance due to issues of measurement error and scope (e.g. Gliem & Gliem, 2003). One often discussed measurement issue is the impact of the item data level (such as nominal, dichotomous, ordinal or interval data) on analysis (cp. e.g. Bortz & Döring, 2006; Moosbrugger & Kelava, 2007; Schumacker & Lomax, 2004). The data level produced by rating scale items, which were used in some of the tests, can be seen as strictly ordinal or treated as interval. For instance, Bortz & Döring (2006, pp. 224; 181-182) show that mean scores are (and can) be calculated for rating scales as long as care is taken while interpreting. Jöreskog and Sörbom (1996, referenced in Schumacker & Lomax, 2004, p. 25) give a value of "[...] 15 distinct scale points" to differentiate between ordinal and continuous variables. However, in geography education studies, scales are often treated as interval at a smaller – and more usually used – number of discrete points such as four to five; e.g. by reporting means (e.g. Hemmer & Hemmer, 2002a; Klein, 2007; Obermaier, 1997, 2005; Uphues, 2007). For achievement data, sometimes only the means of a sum score of the scale are reported (e.g. Hildebrandt, 2006; Lee, 2005), sometimes rubrics and means based on them are used (e.g. Baker, 2002) and sometimes both means of a sum score of a scale and individual item means are used for analysis (e.g. Huynh, 2009).

For greater comparability between different scales, instead of sum scores, average scores will be used in this study (i.e. the sum score divided by the number of items). Additionally, item response theory (IRT) measures will be used for the achievement variables (dichotomous Rasch model) (cp. e.g. Bond & Fox, 2007; Moosbrugger & Kelava, 2007). Rasch estimates are interval scaled (Bond & Fox, 2007, p. 163).

IRT has several other advantages (cp. also summary in Kollar, 2012), including estimating item difficulties and person abilities on the same scale (Bond & Fox, 2007; Moosbrugger & Kelava, 2007) and "[i]nvariance of person estimates and item estimates within the modeled expectations of measurement error over time, across measurement contexts, and so on [...]" (Bond & Fox, 2007, p. 88). Through that, tests or samples can be linked as well as differential item functioning detected (Bond & Fox, 2007). Moreover, through fit indices it is possible "[...] to ascertain whether the assumption of unidimensionality holds up empirically" (Bond & Fox, 2007, p. 35). One issue in that regard is the boundaries in which an item is seen as fitting. For this study, the guidelines of Bond and Fox are followed, namely that t values are seen as acceptable when they "[...] fall between -2.0 and +2.0 with sample sizes between about 30 and 300" (2007, p. 43) and mean square values when they are within the $0.75 < MNSQ < 1.3$ range (ibid. , p. 240). Since fit indices take sample size into account, the larger the sample size, the closer one would expect the MNSQ value to be to 1 and the more t-values are likely to be $> |2|$ (Wu & Adams, 2007). Another metric is item discrimination (see e.g. also Moosbrugger & Kelava, 2007), which according to the testing services site of the University of Wisconsin Oshkosh (n.d.), is unacceptable when below 0.25 and excellent when over 0.4. There are different procedures to produce estimates, such as Plausible Values (PV) and Weighted Likelihood Estimates (WLE). Like in the PISA study, WLE estimates are used (OECD, 2007b, p. 332). Similarly, ability estimates were transformed to a mean of 500 and a standard deviation of 100 for further analysis to ease interpretability.

Regarding sample size for a dichotomous Rasch analysis the minimum seems to be 30, although more than 200 (250) lead to more stable item calibrations as well as a higher confidence and are thus often recommended (e.g. Ewing, Salzberger, & Sinkovics, 2009; Linacre, 1994; Wright & Tennant, 1996). For polytomous scales, "[...] at least 10 observations per category [...]" are needed (Linacre, 1994; cp. also Linacre, 2002, pdf p. 6).

Besides IRT, factor analysis can be used in order to establish unidimensionality (Moosbrugger & Kelava, 2007; UCLA Academic Technology Services Statistical Consulting Group, n.d.). However, while some argue that normal distribution is not essential (see summary in Moysés, 2000, p. 319), dichotomous items can

pose problems (see summary in Garson, 2011), although Robertson et al. (2008) argue that "[t]he use of dichotomous variables can be justified in exploratory approaches [...]" (p. 32). There are special procedures recommended for dichotomous data (see e.g. the discussion on the Mplus page "Binary data and factor analysis", 2000-2011) which seem to be only available in Mplus (see e.g. Durpoix, 2010, p. 334 who references Brown, 2006, p. 388). Another option would be the use of latent class (LC) factor models (Magidson & Vermunt, n.d.; Moosbrugger & Kelava, 2007; Vermunt & Magidson, 2005). However, as a starting point, this study will employ IRT to explore the dimensionality of the items.

Another broadly used measure to estimate scale quality is Cronbach's (1951) alpha. Sijtsma (2009, p. 108) states that "[a]lmost no psychological test or inventory is published without alpha being reported [...]". Despite its ubiquitousness, Cronbach's alpha can be seen critically. Firstly, it is dependent on the number of items used, i.e. in general, at least up to a certain threshold, the more items, the higher the value of Cronbach's alpha (e.g. Moosbrugger & Kelava, 2007; Sijtsma, 2009; UCLA Academic Technology Services Statistical Consulting Group, n.d.). Cortina (1993) argues that "[a]lthough most who use alpha pay lip-service to this fact, it seems to be forgotten when interpreting data. Most recent studies that have used alpha imply that a given level, perhaps greater than 0.70, is adequate or inadequate without comparing it with the number of items in the scale" (p. 101). More specifically, one frequently finds interpretation guidelines such as those of George and Mallery (2003, p. 231, cited in Gliem & Gliem, 2003, p. 87, formatted for better readability): " '

> _ > .9 – Excellent,

> _ > .8 – Good,

> _ > .7 – Acceptable,

> _ > .6 – Questionable,

> _ > .5 – Poor, and

> _ < .5 – Unacceptable' ".

Secondly, alpha "[...] is a function of the extent to which items in a test have high communalities and thus low uniquenesses" (Cortina, 1993, p. 100; see also e.g. discussion in Moosbrugger & Kelava, 2007). Thus, while for instance,

Moosbrugger & Kelava (2007, p. 129) describe the very high values for alpha for established tests measuring global intelligence; Turconi et al. (2003), for their nutritional questionnaire argue that

> "[l]ow Cronbach's alpha in sections H and I could be explained by the items variability; in other words, items of sections H and I cover all aspects of nutrition and food safety knowledge. Nutrition knowledge and food safety knowledge are very complex constructs, and they are made of different domains; therefore, items of sections H and I are quite heterogeneous and this may lead to a low Cronbach's coefficient" (p. 756).

This seems especially relevant to the achievement section of the research reported here, due to the low number of items and the range of systemic thinking skills covered by them. Consequently, also scales with a Cronbach's alpha below 0.70 will be included in the analysis.

Checking for normal distribution will be done with the help of Shapiro-Wilk tests, since sample sizes are below 2000 in all cases (Maths-Statistics-Tutor, 2010).

3.2.2 Connections between variables

Besides test construction, the examination of differences between groups and connections between variables is crucial in order to test the hypotheses of this study. Tests to be used depend on data level. These include

- Chi-square tests for nominal and binary data

- Mann-Whitney-U- (MW) or Wilcoxon-rank-tests for 2 groups and Kruskall-Wallis-tests (KW) for more than 2 groups for ordinal data or continuous data which is not normally distributed and

- t-tests (2 groups) or ANOVAs (more than 2 groups) for continuous data which is normally distributed

(see e.g. Dallal, 2000; Kohlmann & Moock, 2009, p. 97).

For studies using a pre-/posttest design tests of difference for repeated measures are necessary, such as the "Paired *t*-test" for normally distributed data and the "Wilcoxon-matched-pairs-signed-rank test" (W) for non-parametric analysis (see e.g. Dallal, 2000; Kohlmann & Moock, 2009).

The variables included in the AN(C)OVAs need to have a homogenous variance (e.g. Levene's test) and the dependent variable (as well as the covariate) need to be normally distributed (and the dependent variable and the covariate should be correlated linearly) (Schermelleh-Engel & Werner, 2007; Schwab, 2007; Wendorf, 2004). The "[...] dependent variable and covariate are required to be interval level" (Schwab, 2007, p. 1). Variables such as gender or group should be entered as fixed factors (see e.g. García-Granero, 2008) because in fixed factors "[t]he results of the analysis of variance only refer to the investigated values and should *not* be generalized to other conceivable values of the factor" (Schermelleh-Engel & Werner, 2007, p. 1, translated).

ANOVAs and t-tests are, however, quite robust against the violation of the normal distribution assumption for large samples (e.g. BBN Corporation, 1997a, b). Consequently, parametric tests will be applied in addition to non-parametric ones, even for non-normally distributed scales.

AN(C)OVAs, regression analysis, SEMs and other procedures such as LC regression models or latent class growth models can be used to get a more complex view of groups (e.g. Garver, Williams, & Taylor, 2008; Hoe, 2008; Kline, 2000; Magidson & Vermunt, n.d.; Schumacker & Lomax, 2004; Statistical Innovations, 2005; Tabachnick & Fidell, 2007). AN(C)OVAs and regression analyses seem to be much more widely used in geography education studies then the other procedures. In general, AN(C)OVAs and regression analysis are fairly similar (see e.g. Huang, n.d.; Ludford, 2003; Simon, 2010; Wendorf, 2004). Thus, for this report, AN(C)OVAs will be used.

Regarding sample size, for tests of difference, for instance Kohlmann & Moock (2009, p. 100; also: Scanlan, n.d., p. 3) state that the minimum cell size for Chi-square tests is 5. If this condition is not met, other tests, such as "Fisher's exact test" should be employed (Kohlmann & Moock, 2009; Scanlan, n.d.). Moreover, programs such as G*Power 3 show that the power of a t-test is heavily dependent on sample size up to a threshold. The minimum requirement for power is usually 0.80 (Wolske & Higgs, 2010). Calculations in G*Power for t-tests (two-tailed) for independent groups show, for instance, that with equal group size, the total sample size would range between 1302 (alpha=0.05,

power=0.95, d=0.2) and 8 (alpha=0.05, power=0.80, d=2.7). The values are higher if the sample sizes are not equal. For dependent means t-tests (two-tailed), the values range between 327 (alpha=0.05, power=0.95, d=0.2) and 4 (alpha=0.05, power=0.80, d=2.7). For ANCOVAs, Schwab (2007, p. 1) states "[...] a minimum sample size requirement of 5 cases per cell [...]".

Although other p-values (such as 0.10 or 0.01) are sometimes used, this study will follow the widely used p=0.05 as a cut-off-value for statistical significance (cp. e.g. Dallal, 2003; Kohlmann & Moock, 2009; Wolske & Higgs, 2010). The p-value, however, is dependent on sample size and thus effect sizes are often seen as (more) useful to report (e.g. Levine & Hullett, 2002; Rost, 2005a; Valentine & Cooper, 2003; Wolske & Higgs, 2010).

Effect sizes enable comparing differences across studies as well as helping in interpreting the importance. Cohen's d, a measure for effect size used for t-tests, can be interpreted as small when 0.2, medium when 0.5 and large when 0.8 (Cohen, 1988). Ferguson (2009, p. 533) states that 0.41 is the "[...] recommended minimum effect size representing a "practically" significant effect for social science data" (RMPE) for using d, a moderate effect would be 1.15 and a strong effect 2.70. Hattie (2009, p. 17) argues that an "[...] effect size of 0.40 sets a level where the effects of innovation enhance achievement in such a way that we can notice real-world differences [...]". However, taking into consideration the very short duration of the treatment, an effect size of 0.4 seems comparatively high. Consequently, also small effects (0.2) are reported.

For effect size measures such as (adjusted) R^2 or η^2 Ferguson (2009, p. 533) recommends 0.04 as RMPE and interprets 0.25 as a moderate and 0.64 as a strong effect. Although SPSS only reports the partial η^2 values, η^2 could be computed from the SPSS output (see e.g. "Example Calculation of Eta-squared from SPSS Output," for the procedure). Thereby, "[...] eta squared is often interpreted in terms of the percentage of variance accounted for by a variable or model" (Levine & Hullett, 2002, p. 619). Moreover, eta squared has other positive characteristics such as being "[...] useful for subsequent meta-analyses [...]" and "[...] the property that the effects for all components of variation (including error) will sum to 1.00" (ibid.). However, partial eta square has

advantages when comparing "[...] an effect of an identical manipulation across studies with different designs [...]" (Levine & Hullett, 2002, p. 620).

For this study partial η^2 values will be reported in case of AN(C)OVAs and d values will be used for t-tests.

Another way to analyze data is by looking for correlations. Correlations describe the strength of a (linear) association as well as the direction (positive, negative) of this association. There are different statistics that can be used. For instance, Pearson's product-moment-correlation coefficient (r) is used for examining two interval scaled variables, Spearman's rank correlation (r_s) (or Kendall's τ) for ordinal scaled variables as well as as a non-parametric test for interval scaled data, a point biserial correlation (r_{pb}) for one dichotomous and one interval scaled variable, and a coefficient of contingency for associations involving one or two nominal variables (see e.g. Bortz & Döring, 2006, p. 508; MEI, 2007; Scanlan, n.d.; Schumacker & Lomax, 2004; Valentine & Cooper, 2003). For the later, measures include Phi (Φ) for 2*2 tables and Cramer's V for larger tables ("Nominal measures of correlation: Phi, the Contingency Coefficient, and Cramer's V," n.d.; Scanlan, n.d.; Schumacker & Lomax, 2004). Ferguson (2009, p. 533) states for both Φ and r that the RMPE is 0.2, a moderate effect would be 0.5 and a strong effect 0.8.

For correlation analysis, power is dependent on sample size. Ahmed & Capretz (2008, p. 1104, according to them based on Cohen's 1988 work) describe that the minimal sample size requirements for Spearman correlations are higher than for Pearson correlations, and for a power of 0.8 vary between 618 and 12 for Pearson correlation coefficients of 0.1 and 0.7, respectively, and 733 and 15 for Spearman correlation coefficients of 0.1 and 0.7. Schuhmacker & Lomax (2004) point to the additional concern that the sample should not be too homogeneous, i.e. "[...] that there must be enough variation in scores [...]" (p. 40).

One problem in educational research, especially on the school level, is that participants usually belong to a particular group (a class). As such, they can be described in a hierarchal order, e.g. level one=students, level two=class, level three=school and so on, or also different points of assessment for one student (Ditton, 1998). Both variables that are commonly included in the analysis as at-

tributes of the level one student (e.g. stream) and those that are discussed as confounding variables (e.g. teacher characteristics) actually belong to the level two and are the same for all members of the class (Ditton, 1998). This leads to problems with statistical procedures such as significance tests and regression analyses (for details see Ditton, 1998, pp. 32-34). Consequently, Ditton (1998) argues that the "[a]nalysis of hierarchical data under ignorance of their multi level structure can lead to grave mistakes and possibly even be totally useless!" (p. 13, translated, in original italics). Similarly, Hox (1998) states that "[a]ssuming a common class size of 25 pupils, and a typical intraclass correlation for school effects of rho=0.10, we calculate an operating alpha level of 0.29 for tests performed at a nominal alpha level of 0.05! Clearly, in such situations **not** adjusting for clustered data produces totally misleading significance tests" (p. 148). Ditton (1998) argues that aggregating the data and afterwards only analyzing the level 2 is also not to be recommended, especially due to the information loss. Thus, Ditton (1998), Hox (1998) and others recommend utilizing multi-level analysis methods such as hierarchal linear modeling (HLM) for the analysis of such data.

Until now, however, HLM seem to be rarely used in geography education research. One possible reason is the great demands on sample size. In general, minimum sample size recommendations seem to vary between about 10 and 30 for level 1 (students) and 30 and 300 for level 2 (classes) (e.g. Ditton, 1998; Hox, 1998; Langer, 2005; Maas & Hox, 2005), whereby Langer (2005), referencing the work of Kreft (1996), states that "[...] in case of doubt rather more level-2- as level-1 units [...]" should be used (p. 19, translated; see also Hox, 1998). Study 5 is very close to the lowest limit requirements with 44 classes, with nine (one class) to 29 students, when analyzing the whole sample. However, separate analyses of the map and GIS group are not feasible.

At the beginning an intercept only model is calculated (based on Hartig, 2005). Based on UCLA Statistical Consulting Group (n.d.) the intraclass correlation, i.e. "[...] the portion of the total variance that occurs between [...]" classes, can be calculated by dividing the variance component value of u0 by the sum of the variance component of u0 and R (see also SSI Central, n.d.-b). Based on the intercept only model, predictors are added individually (based on SSI Central,

n.d.-b; UCLA Statistical Consulting Group, n.d.). The variance in outcome over classes controlling for a level 2 predictor can be calculated by the difference of u0 (unconditional) and u0 (model with added variables) divided by u0 (unconditional) (based on SSI Central, n.d.-b; UCLA Statistical Consulting Group, n.d.). Next, looking at the level one predictors individually, the proportion of variance explained at level 1 can be calculated (according to UCLA Statistical Consulting Group, n.d.) (within class variance) as the difference between R (unconditional) and R (model with added variables) divided by R (unconditional).

However, using the "proportion variance statistics" can be problematic, including "[...] interpretation can cause confusion and estimation anomalies can occur" (SSI Central, n.d.-c, p. 137). With reference to Raudenbush & Bryk (2002, pp. 149-152), SSI Central (n.d.-c, p. 137) state "[...] that the variance explained in a level-2 parameters [...] is conditional on a fixed level-1 specification, and that the variance reduction statistics are only interpretable for models with the same level-1 model." Consequently, level 1 variance explained is compared to the intercept only model, while level 2 variance explained is compared to the model with all the respectively used level 1, but without the level 2 variable.

Sex, migration and family language background, group/materials and stream are entered as uncentered and pre-score, age as well as pre-spatial thinking score (and total language) as group mean centered (based on Graham, 2007). Because the number of classes is not large, restricted maximum likelihood estimation should be used (SSI Central, n.d.-a). Moreover, only main effects are examined, not cross-level-effects, since the number of classes is below 50 (Radisch, 2010). Data is exported in ASCII format from SPSS and imported into HLM 6.08 and the statistics checked.

3.2.3 Software used for analyses

The data analysis is done in Conquest and HLM 6.08 as well as successive versions of PASW Statistics/SPSS. Effect sizes are calculated with the help of the website provided by Ellis (2009) for independent sample t-tests or G*Power 3 (http://www.gpower.hhu.de/) for paired sample t-tests. Power analysis is also done in G*Power 3.

3.3 Sample procedures

Studies 1 to 4 were done only in individual institutions in Baden-Wurttemberg in spring and summer 2008 as well as the winter term 2008/2009. Study 5 was conducted in spring and summer 2009 in three federal states, namely Baden-Wurttemberg, Schleswig-Holstein and Saxony.

Originally it was planned to conduct the study in five different states, for which official approval had been sought in autumn 2008. However, only three answered in time. Approval was granted for mid and high stream schools in the administrative district Karlsruhe (Baden-Wurttemberg) and for a list of schools in Schleswig-Holstein. Saxony did not permit the study for state schools, but to conduct it in private schools was possible. Consequently, the Saxon sample is very small.

The original five states were selected based on the thematic fit of the treatment unit to the curricula, overall score in PISA reading literacy 2003 and score in PISA mathematics literacy scale "space and form" and status of GIS integration into the curriculum. Thematic fit was of great importance, since a learning unit without explicit curriculum support would have made it difficult both to get approval and to, afterwards, get teachers to participate in the study. Reading literacy is associated not just with gaining information from continuous texts, but also from media such as maps (e.g. Hüttermann, 2004; Lenz, 2003; OECD, 2006a). Mathematical literacy, especially with regard to the scale "space and form" is often said to be related to spatial thinking in geographical space by non-geographers. From the five federal states, two had above average scores both for the reading literacy and the scale "space and form" of the mathematical literacy (Baden-Wurttemberg and Saxony), two had below average scores (Rhineland-Palatinate and Mecklenburg-Western Pomerania), and one had average (maths) to below average (reading) scores (Schleswig-Holstein) (Prenzel, et al., 2005). Consequently, most of the students participating came from above average performing states.

To recruit participants, letters were written to schools in Baden-Wurttemberg, Schleswig-Holstein and Saxony. Additionally, personal contacts were employed. As a follow-up, e-mails were sent to schools that had not yet responded. Some schools that had originally expressed interest did not partici-

pate in the end, due to various reasons; participated, but did not send back the data; or participated with a number of classes that was different than originally indicated, contributing to some of the imbalance in the data.

In all studies, the tests were provided, as well as the work sheets and other materials. In study 5, teachers were also e.g. provided with a short summary of information about tourism in Kenya as well as an overview over the study/ organizational information. In study 5, on a limited basis, classes in Baden-Wurttemberg that did not have the required atlas could borrow it for the duration of the study. Moreover, in study 5, teachers interested in participation in the study were additionally offered a free and non-obligatory professional development course giving an introduction to GIS. Courses were held at the Department of Geography at the Heidelberg University of Education, at three individual schools, and at the University of Kiel. The training generally had both theoretical and hands-on parts, and a variety of topics including what GIS is, how it can be used, claims vs. state-of-research, the role of GIS in curricula and practical hints and links. Hands-on practice focused mainly on tasks from the study and the Diercke WebGIS. Moreover, before and during the study, teachers could get support via telephone and e-mail if needed. In three schools, the author was present for part of the study. In study 5, when the study was finished, the materials were returned either through being collected by the author, through the teachers sending a postal package, or through the teachers coming and bringing back the materials in person.

For pre-/posttest studies, procedures for matching the tests are necessary. While unique codes were provided on the pretest of study 3 which the students had to note down and put onto the posttest and the materials, this procedure was advised as too insecure for younger students and a treatment stretching out over several days. Consequently, a code was used (see ch. 4).

4 Describing the samples

Several common variables are used not only to test the hypotheses but also to describe the samples in the test. All variables were originally in German and are translated here where necessary. Due to test time considerations, not all variables were included in each of the studies.

4.1 Variables

Several demographic variables that can be used to describe the sample were included in the questionnaires. After a description of the variables used, in line with the focus of this dissertation, sample characteristic descriptions for studies 1 to 4 will be only brief, not covering all included variables, while for study 5 all included variables will be described.

4.1.1 Basic demographic variables

One of the basic demographic variables important for the hypotheses is sex. In study 1, still a phrase was used (similar to Sommer, 2005). In the later studies it was shortened to "sex:" (similar to Uphues, 2007) with students being able to check male or female, which seemed to work with the students as well. Similarly, the age question was shortened from study 1 to the later studies. In study 2 the students also had to check the grade level they were in, because it was conducted in more than one level within one school. In study 5 this variable was also included, while in study 4 the corresponding question asked for the semester the students were in. In study 1 (as a phrase, similar to that of sex, above), as well as in studies 2, 3 and 5 (shortened version) students were also asked to state their stream. In study 4, it was asked which role geography played in their studies (e.g. major subject at the University of Education, *Diplom* at the University, four choices plus blank plus "not my subject") instead.

4.1.2 Language background

The language background was more difficult to assess. Study 1 included two questions: "With my family I speak these languages: ____" and "Besides, I also know/am learning these languages: ____". This was preferred over the PISA-

wording ("What language do you speak at home most of the time?", p. 11) and coding (only one choice) (OECD, 2000) to get a more complete view of the students' language background. This seemed important because, for instance, families with migration background often speak more than one language with their children and students might achieve a near-native level in languages they do not speak with their family. Looking at the answers, the phrasing seemed to work, although there were some cases that needed to be decided upon. For instance, in study 1, one student wrote "English/American" as family languages. Although 'English' was not written after American in contrast to study 5, the student probably referred to American English. A second example was one student that clearly stated that English is only spoken at home when practicing for a test. For this student, English was treated as other language. German was added as a learned language for students who seemed to have forgotten it, since the test was in German.

Since cell sizes would be very small, especially for rarer languages, different possibilities of grouping were tried out, such as the most frequent combinations plus "other" or a grouping by the second category in the *ethnologue* database (SIL International). For instance, German and English can be classified as Germanic and French and Spanish as Italic. Family languages can also be categorized as German, German and one other, German and two or more others or only other languages. Even more condensed, it could be classified as "German only" and "other". The other languages naturally included mainly learners of "classical" school languages such as English and French, but also a small minority of learners of other languages (e.g. Esperanto) and could be classified accordingly. While this does make it possible to sketch a rough picture, it has to be kept in mind that the phrasing of the question neither allows for statements regarding the quantity of use nor regarding the abilities in each of the languages.

Consequently, in study 2, it was tried to get a feedback on how well the students knew the languages. The item was consequently formatted as a table with the headings "I think in these languages: ___" and "In these languages I can" with the categories "only a few words and sentences", "communicate in some basic everyday situations", "communicate in most everyday situations", "communicate also about geographic topics", "communicate fluently about (nearly) all topics". The

categories were adapted from the Common European Framework of References for Languages (Council of Europe, 2001), whose even short format level-description still requires considerable reading time. According to the teacher feedback and the student answers, this question was fairly difficult to understand for the students. For instance, there was a high number of missing values. Looking at selected competency levels, several students did not specify levels at all and several checked more than one level (e.g. the highest three). Moreover, two students did not specify the highest level for German, one of which was born in Germany as where both of the parents and one who was born in Germany but both parents were born abroad. Additionally, one student specified the highest level in English, and two more specified the highest level in English among several levels in English (all three without migration background). This level is far beyond the current curricular aims for that age group and stream (MKBW, 2004b). Furthermore, if English (only where it is not home language) is classified so that the highest level specified is counted, there is no significant difference between the two grades, contrary to expectations (crosstabs, Cramer's V=0.24, p=n.s., n=65). However, a high percentage of cells do not fulfill the minimum requirement. These aspects support the opinion that the item is problematic.

Studies 3 and 4 did not include language background variables. Due to the problems with competency self-assessment, in study 5 the same phrasing as in study 1 was used.

4.1.3 Migration background

The migration background was not included in studies 1, 3 and 4. In study 2, two different items were tried. The first item focused on migration background and asked for the birth countries of the student as well as the student's father and mother (e.g. "I am born in this country: __"). While in PISA 2000, the students were only asked for "Country of test" or "Another country" (OECD, 2000, p. 10), in PISA 2003 several options were specified (OECD, 2003a, p. 9). Instead of specifying the countries of the most common immigrant nations, however, the question was given in an open answer format. The second item was an attempt to employ the concept of ethnic groups as an addition to – or possibly a replacement of – asking for the birth countries. The attempt was contemplated especially based on re-

search results in the USA, where the concept is commonly employed and significant differences in educational achievement have been documented (for geography education, see e.g. overview in Bednarz, et al., 2013; for a general overview see e.g. Kao & Thompson, 2003). However – maybe because of it being fairly uncommon in Germany in contrast to the USA – it seemed to be not understood by the students and did not produce any usable results. In study 5, the same phrasing as in item one for study 2 was used (birth countries).

Since many individual countries are only represented in very small numbers, especially in the smaller studies, they could be categorized by area, e.g. into "EU and Switzerland excluding Germany", "Turkey" and "other areas" or "former major *Gastarbeiter* (indentured laborer) countries" and "other countries". However, even then cell sizes are small.

Another categorization would be into "student born abroad", "student born in Germany, both parents born abroad", "student born in Germany, only one parent born abroad", and "student born in Germany, both parents born in Germany". Very few students are born abroad but have one or both parents born in Germany. More generally, students could be categorized into "migration background" or "no migration background".

4.2 Study 1

The first study was conducted with grade seven students from one comprehensive school (*n*=102) in spring 2008. There are 49 boys and 42 girls in the sample, with eleven coded missing. The age ranges from twelve to 14 (seven coded missing, *M*=12.8, *SD*=0.52, *Mdn*=13, *n*=95). There are 47 students stating to be in the high stream and 51 in the middle stream, with four coded missing.

4.3 Study 2

The second study was conducted with grade seven and grade nine students from one mid stream school (*n*=98) in spring/summer 2008. There are 42 boys and 55 girls in the sample, with one coded missing. The age ranges from twelve to 17 (*M*=14.7, *SD*=1.1, *Mdn*=15, *n*=97). In class seven, the mean age is

13.2 (*SD*=0.65, *Mdn*=13, *n*=26, range twelve to 14) and in class nine 15.2 (*SD*=0.62, *Mdn*=15, *n*=71, range 14 to 17).

4.4 Study 3

The study was conducted with grade twelve students of one high stream school in summer 2008, shortly before the summer holidays, which could be seen as an unusual setting. Students were assigned to either the atlas or the GIS group by the teacher and worked on the unit during one day, taking the pretest in the morning and the posttest at noon. There are 64 students in the sample, however, since one posttest went missing, all analyses are done with *n*=63.

Only sex and age were included as demographic variables, as in light of test time considerations the improvement of the spatial and systemic thinking items had higher priority than the development of the language skills section. There are 24 boys and 38 girls in the sample, with one coded missing. The mean age is 18.7 (*SD*=0.84, *Mdn*=18, *n*=62, range 18 to 21).

There are 32 students in the map and 31 students in the GIS group, with no significant difference with regard to age (MW, *p*=n.s., *n*=62) or sex (crosstabs, Φ=0.07, *p*=n.s., *n*=62) between the two groups.

4.5 Study 4

The study was conducted in the winter term 2008/2009 with students from two universities. A total of 190 students participated. The main aim is to further explore the difficulty of the spatial thinking items as well as to test a simplified version of the spatial thinking items. Moreover, the study gives first exploratory insight into the spatial thinking skills of students of geography and geography education.

Of the sample, 45 are in group A (four items with the graphics used earlier), the rest in group B (four items with the graphics used earlier and seven items with simplified graphics). The age ranges from 19 to 31 (*M*=22.8, *SD*=2.6, *Mdn*=22, one missing, *n*=189). The semester ranges from one to 13 (*M*=3.7, *SD*=3.3, *Mdn*=3, three missing, *n*=187). There are 83 males and 107 females in the sample.

The item eliciting the students' background with regard to their study of geography did not fully cover the diversity found in the sample in its pre-defined an-

swers, and students were sometimes not very specific in their answer in the "other"-field. Consequently, there is a fairly high percentage of students that cannot be fully classified. There are two reasonable groupings (see Tab. 2, *n*=190, % of total sample). The other group for grouping one includes students that only specified "minor" or "major", without stating whether this referred to a teacher or a non-teacher course, a *Magister* student, a student specifying more than one course etc. For grouping two, additionally the students specifying minor (bachelor/*Diplom*) or not giving further specification (e.g. only, "bachelor" or "state exam") are included in this group. Thereby, specifications such as "major state exam" were classified as belonging to the high stream, since "major (PH)" was one of the provided answers.

grouping 1	*n*	%	grouping 2	*n*	%
(a) teacher	83	43.7	**(i) teacher major (primary, lower/middle stream, special needs)**	43	22.6
(b) bachelor/*Diplom*	67	35.3	**(ii) teacher major (high stream)**	26	13.7
			(iii) bachelor/*Diplom* major	57	30.0
			(iv) teacher minor (primary and all streams)	12	6.3
(c) other/not further specified	32	16.8	(v) other/not further specified	44	23.2
(d) not my subject	3	1.6	(vi) not my subject	3	1.6
missing	5	2.6	missing	5	2.6

Tab. 2: Sample description by study of geography background

4.6 Study 5

Study 5 was conducted in three federal states in spring and summer 2009, with the data being received back till September 2009. The pre- and posttest as well as the materials were received and matched by class. The clear seeming cases were matched and the code sheets destroyed as advised by data security. The code used comprised the initials of the parents given names and their birth months. This code proved difficult for many students and – although similar codes are used in other studies (e.g. Hildebrandt, 2006, p. 126; Kramer, 2005, p. 193) seems to not be advisable for this age group. Moreover, the codes alone proved to be problematic in some cases. One teacher noted that there were twins in his class, but that the pre- and posttest were unambiguously matchable based on their handwriting. In consultation with the first advi-

sor as well as others, this procedure was also applied to a few other cases of same codes as well as some were the codes did not completely match (e.g. different birth months), with some being matchable and few being judged as too doubtful and thus not matched. Difficult cases were shown to several persons to base the matching – or not matching – not on a sole opinion.

The matching difficulties also exposed inherent weaknesses in the coding system. Firstly, in the unlikely event that in one class there were two students with the same code (e.g. twins) and one handed in only the pretest, and the other only the posttest they would be considered one student. Secondly, some students might have been judged as two different students, e.g. because of too large differences in the code, even if they were in reality only one.

When matched, the pre- and posttest were each labelled with a student number, a school number and a class number. In some cases, class membership within one school was unclear, especially when the materials were sent back sorted only by group, not class. However, based on the times, these could be usually differentiated.

After the coding, exclusion criteria had to be decided upon, which were based on the research questions as well as statistical considerations (e.g. excluding grade eleven students because of having no suitable comparison group). These were, in order:

(a) students not having both tests

(b) students being in class eleven (one class, 18 students). One student (class ID 24) specified twelfth grade, however, based on teacher feedback was classified as grade seven. Two students (class ID 2 and 14) did leave the class question blank and were classified as grade seven.

(c) students not completing the test(s) individually, irrespective of whether they used color differentiation or not; because even students who color coded their partly different answers might have been influenced by each other and thus not comparable to the others.

(d) students not answering the question with which medium they worked with or contrary to expectations or checking both, as at least two teachers had some students work with a medium they did not have the respective version

of the materials for. Consequently, these students were excluded as well, as they would be not comparable to others in either group. Students were excluded irrespective of whether the school had access to the materials they specified (e.g. due to participating with more than one class) or not, solely based on coded class membership.

(e) students not having coded materials and coming from schools that had access to both sorts of materials, as it appears that at least one of the remaining students – in a school that had access to both GIS and map materials – had been coded as being in one class and provided the expected answer to the question which medium was worked with, but the matching code appeared on the materials for the other medium.

(f) students not taking the questionnaire seriously at all (e.g. adding *"Baumschule"* (tree nursery) or *"Affenschule"* (monkey school) under "other" for school type, and/or checking that he/she is both male and female.

The 932 students that passed these criteria are from 18 schools in Baden-Württemberg (28 classes), six schools in Schleswig-Holstein (13 classes) and two schools in Saxony (three classes). The resulting class sizes vary between nine and 29.
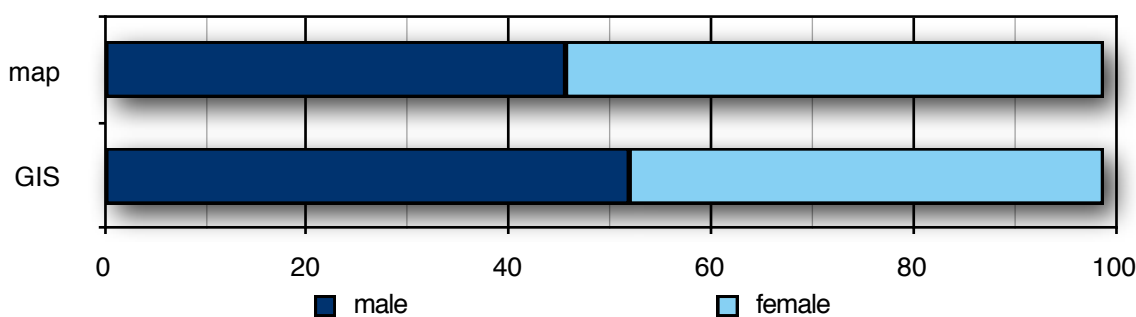
Fig. 8: Sample description by group and sex
        (% of group)

Of the sample, 463 are male, 458 female, with eleven coded missing (Fig. 8). When the one student who forgot to specify the stream is given the class' stream, there are 622 high stream students and 310 mid stream students (Fig. 9). There is no significant difference with regard to sex between the two streams (crosstabs, Φ=-0.02, *p*=n.s., *n*=921). There are 329 students in the map group and 603 students in the GIS group. There is no significant differ-

ence between the two groups with regard to stream (crosstabs, Φ=-0.01, *p*=n.s., *n*=932) or sex (crosstabs, Φ=0.06, *p*=0.07, *n*=921).
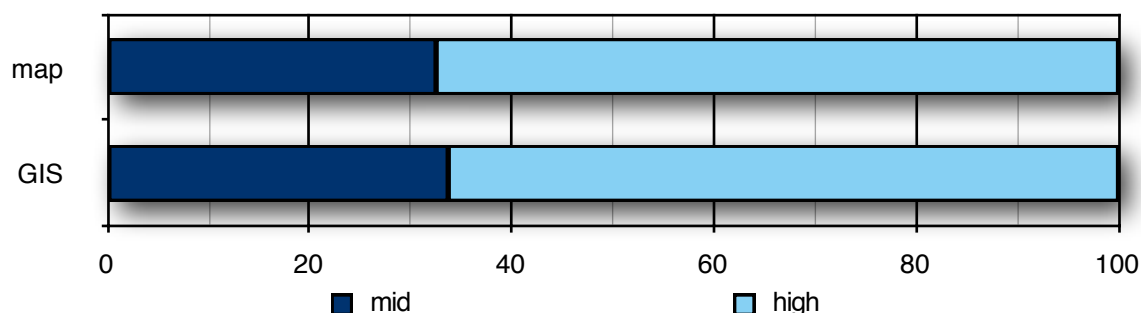
Fig. 9: Sample description by group and stream
(% of group)

There are 236 students in the map 'old' group, 93 in the map 'new', 441 in the WebGIS "old" and 133 in the WebGIS "new" (Fig. 10). ArcGIS was only used by a small group of high stream students (*n*=29). Additionally, in the ArcGIS group, part of the sample did the posttest considerably later than the unit.
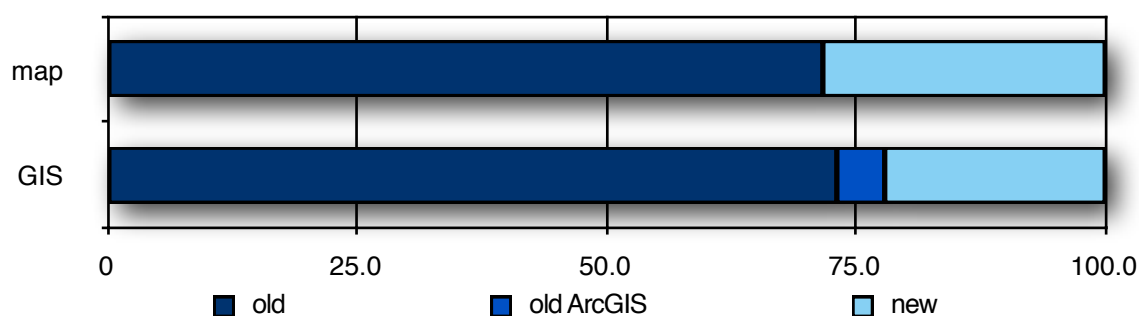
Fig. 10: Sample description by group and type of materials
(% of group)

Within both the GIS and the map group, the different material type sub-groups are significantly different with regard to stream (crosstabs, map Φ=-0.40, *p*<0.001, *n*=329, GIS Cramer's V=0.38, *p*<0.001, *n*=603). While 84.2% of the map group high stream students worked with 'old' materials, only 45.8% of the map group mid stream students did. Similarly, in the GIS group 81.5% of the high stream students worked with 'old' WebGIS materials, 7.3% with 'old' ArcGIS materials and 11.3% with 'new' WebGIS materials, while of the middle stream GIS students, 56.7% worked with 'old' and 43.3% with 'new' WebGIS materials.

With regard to age, there is one student who entered "7". Since this is the same as the grade level which is asked for just above, this might be accidentally entered, as it would be very unlikely that there is such a young student in

grade seven. Without this student, the mean age is 12.8 (*SD*=0.59, *Mdn*=13, range eleven to 16, four missing). The age groups eleven, 15 and 16 have very small cell sizes. There is a significant difference between the map and GIS students with regard to age (MW, *p*=0.03, *n*=927), with the map group having a mean of 12.7 (*SD*=0.58, *Mdn*=13, range eleven to 15, *n*=328) and the GIS group of 12.8 (*SD*=0.59, *Mdn*=13, range eleven to 16, *n*=599) (Fig. 11).
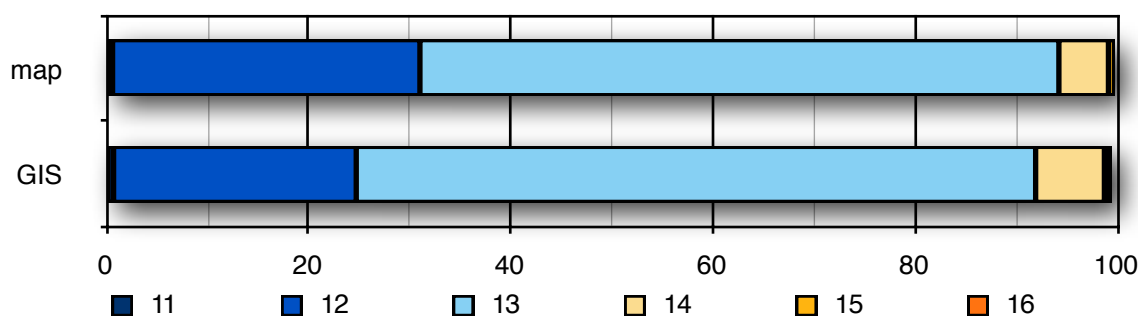
Fig. 11: Sample description by group and age
(% of group)

Overall, students are treated as having a migration background if they themselves and/or one or both of the parents were not born in Germany. For the migration background the former GDR (German Democratic Republic) is also classified as "Germany". There are some students that are somewhat difficult to classify, e.g. one specifying different countries for "1. father/2. father", with one being Germany and the other not; or one student specifying two countries for the mother. Moreover, there are students specifying "Silesia", which cannot be fully assigned to one present-day country. As these are very few cases, they are excluded for the respective analyses, as are students not specifying one or both parents' birth countries. Other special cases include one student's father born abroad, but according to the student holding German citizenship (treated as abroad) and one student's father born in Germany but being half of another country's origin (treated as Germany). Moreover, there are students answering simply "yes" or "no" in all or some of the questions. This could signify that they understand "this country" to mean "Germany". These students are counted for migration background but not for migration background by area. Furthermore, it is notable that two of the students who are not born in Germany themselves, have one or more parents born in Germany.

Applying these criteria, 155 students can be classified as having a migration background, and 770 as not having one (seven coded missing, *n*=932). There is

no significant difference between map and GIS group (crosstabs, Φ=-0.06, *p*=0.05, *n*=925, Fig. 12). Looking at the migration background (in total 65 for map and 90 for GIS group) by area cell sizes are small, the largest groups coming from the EU/Switzerland (24 in the map and 37 in the GIS group) and the countries of the former SU (17 in the map and eleven in the GIS group). Due to the cell sizes, as well as most cells' diversity, it does not seem advisable to conduct analyses by area. Since in total only 20 students are first generation migrants, it seems not advisable to split the migration background by generation.
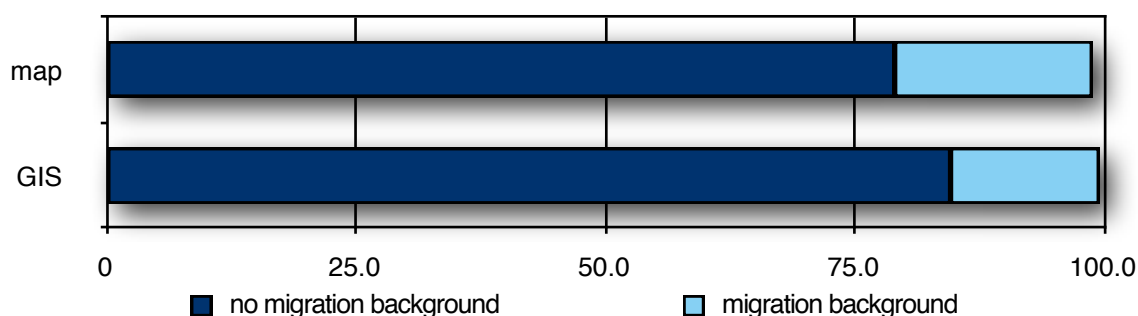
Fig. 12: Sample description by group and migration background (% of group)

Looking at the language background, German was added when students forgot it, since the test was in German, and English was added based on the replies of the teachers that it was compulsory in their school. Languages to be learned in the future (e.g. "next year Spanish"), computer languages (e.g. "basic") or languages such as "a secret language I invented myself" were not entered. A problem with language coding is the differentiation between what is an independent language and what is only a national variety or dialect. While some like American English or Swiss German are fairly easy (recorded as English and German respectively), the status of others is debatable. Some languages, such as Low German have been included in the European Charter for Regional or Minority Languages (Council of Europe, 2002), and thus have been coded as an independent language for this study. Other languages, such as Silesian do have an ISO language code (see the listing at www.sil.org), but seem not to be officially recognized as minority language yet (Council of Europe, 2010). However, Silesian may refer to a Polish or a German dialect (compare the listings for www.ethnologue.com/show_language.asp?code=sli with those for code=szl). Consequently, it is not included as a language. "A Swiss dialect" is also not included as a separate language.

The number of languages ranges between two to six (*M*=3.0, *SD*=0.60, *Mdn*=3, *n*=932, Fig. 13). There is no significant difference between map and GIS group (crosstabs, Φ=0.09, *p*=n.s., *n*=932).
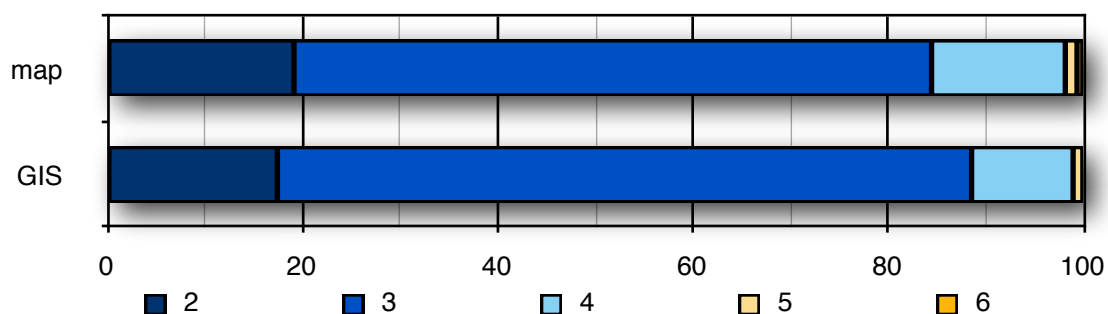
Fig. 13: Sample description by group and number of languages (total) (% of group)

Most students (*n*=792) specify speaking only German with their family. Both German and other language(s) are specified by 129 students (one other *n*=121, two others *n*=8). Only eleven speak only another language with their family. There is no significant difference between map and GIS group (crosstabs, Φ=0.08, *p*=n.s., *n*=932, Fig. 14).
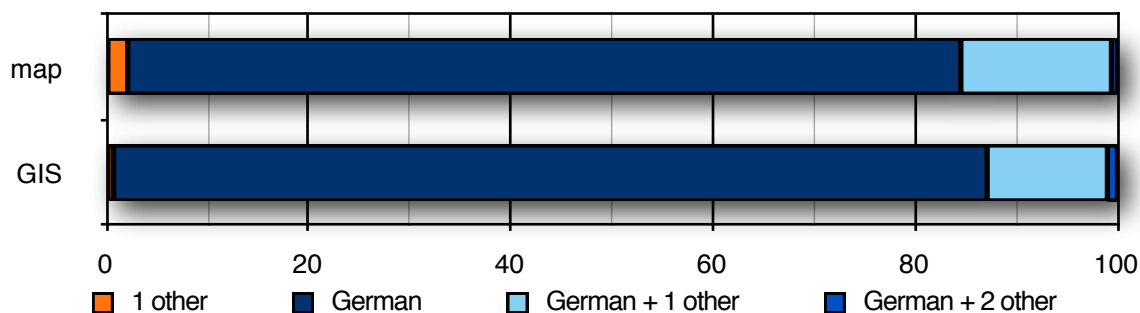
Fig. 14: Sample description by group and family language coding (% of group)

4.7 Limitations due to the achieved samples

The achieved samples lead to several limitations. The first limitation relates to sample sizes, which are comparable to or fairly large for geography education studies (see ch. 2) but small in comparison to large scale studies. This especially effects certain cell sizes. Moreover, it also leads to no interactions being explorable by HLMs. The second limitation relates to the imbalance (e.g. between the two different streams) in the sample and that students stem from only three states. This poses constraints on the types of analyses that can be done and the interpretation of the results.

5 Describing the instruments

"The field of geography education is sadly lacking in empirical data
that might inform and underpin decisions [...]" Downs (1994, p. 57).

Downs' assessment is still sadly valid for the situation of geography education nearly two decades later. His assessment can also be applied to the German situation, since little seemed to exist in terms of standard and empirically validated German test instruments for assessing the achievement of geography education's guiding objective. This posed a problem for this dissertation. There were three options: using one of the few existing German assessment instruments for systemic thinking, translating and using an existing international systems thinking assessment instrument or developing an own assessment instrument.

Regarding the first option, many of the German instruments found either deal with a fictitious situation such as the Hilu and Mori tribes (e.g. Klieme & Maichle, 1991, 1994; Maierhofer, 2001; Ossimitz, 1996, 2000) not stemming from a strictly geography education project or have been developed for other subject domains such as biology (e.g. Sommer, 2005). Due to the short possible duration of the treatment it would be doubtful whether any changes would be measurable on a test whose context differs so much from that of the treatment. Moreover, despite domain boundaries in general being fuzzy, this study aims at exploring the effects of GIS use on systemic thinking in geography, not domain unspecific or biologic systemic thinking. Furthermore, neither the HEIGIS instrument (see e.g. Siegmund, et al., 2012; Viehrig, et al., 2011; Viehrig, et al., 2012) nor the instrument by Rempfler and Uphues (see e.g. Rempfler & Uphues, 2012) had already been developed at that time.

Regarding the second option, besides tests from other domains, several assessment instruments in the geosciences had been published. These often deal with topics such as the water or the rock cycle (e.g. Ben-Zvi Assaraf & Orion, 2005a, b; Kali, et al., 2003; Sibley, et al., 2007). These topics are not part of the geography curriculum for Baden-Wurttemberg for the high stream for grades seven/eight, but the rock cycle is a topic for grades nine/ten (MKBW, 2004a).

However, helping students "explain the development of rocks as a cyclic process" (MKBW, 2004a, p. 242, translated) and thus find the answers to questions such as those used in the assessment instrument developed by Kali et al. (2003) does not lend itself easily to an exploration with GIS. Furthermore, Web-based GIS applications specifically developed for use in schools (e.g. http://diercke. webgis-server.de, www.webgis-schule.de, http://webgis.bildung-rp.de, www.sn. schule.de/~gis/) tended to not include geologic themes. The notable exception was the Web-based GIS published by Klett (www.klett.de/sixcms/list.php?pa ge=lehrwerk_extra&titelfamilie=Haack%20Weltatlas,%20Haack%20Weltatlas %20SI&extra=Klett-GIS%20Projekte) which included a geologic map of the federal state of Saxony. Moreover, study results show that in Germany "rocks" belongs to the least interesting topics for students (Bayrhuber, et al., 2002; Hemmer, et al., 2007).

Consequently, the third option was chosen. It was decided to develop an own assessment instrument, based on existing works, but using the context 'tourism in Kenya'. The items were formulated after perusal of and partly adapted from existing studies, both in geoscience and in other contexts. During the test development different versions were discussed with experts, experienced teachers, teacher students and individual adult and secondary school student volunteers. Sub-sets of items (in German) were tested with small samples in studies 1 to 3 (spring and summer 2008), and with two somewhat larger opportunity samples in studies 4 (winter term 2008/2009, spatial thinking only) and 5 (second half of the school year 2008/2009). This report will focus on a selection of variables, i.e. the systemic and spatial thinking aspects.

5.1 Definition and model selection

As 'system' is such a widely used concept, no uniform definition of systemic thinking exists. For this dissertation, based on previous works on systemic thinking and the general competency definition of Rychen and Salganik (2003), geographic system competency (GSC) has been defined "[...] as the knowledge, the cognitive and practical skills, as well as attitudes, values and motivations so as to be able to analyze and comprehend geographical systems in specific contexts" (Viehrig, Volz, & Siegmund, 2008, p. 429). While often in-

cluded in other definitions (see ch. 2), skills such as acting towards systems have been excluded in order to focus the research.

dimensions	components				
content area	population	vulcanoes	Kenya tourism	...	
scale	local	regional	national	international	global
context familiarity	I low				
	II high				
task complexity	I identify elements				
	II identify relationships between elements				
	III place elements into a network of relationships				
	IV perceive hidden components and development of a system within time				
attitudes/ motivations/interest	I low				
	II high				

Fig. 15: Model of GSC
          (based on Ben-Zvi Assaraf & Orion, 2005a; DGfG, 2007a; Rychen & Salganik, 2003)

Of the many proposals for a structure of GSC (see ch. 2), a selection for a model to serve as foundation for assessment instrument development had to be made. This dissertation uses a model largely based on that of Ben-Zvi Assaraf and Orion (2005a), whose model constitutes the components of the dimension "task complexity" in a slightly adapted form. Their model was chosen because a GSC-model should be as general as possible so as to be able to be used with a variety of curricular topics, yet specific and simple enough to be measurable. Moreover, it is a model developed both on a theoretical and empirical foundation, with the empirical foundation being drawn from the geosciences context. Furthermore, the German National Educational Standards for Geography (DGfG, 2007a) and the work of Einhaus (2007), as well as the GSC definition were especially important for the generation of the model's remaining dimensions (Fig. 15). This model differs from the later HEIGIS model (Siegmund, et al., 2012; Viehrig, et al., 2011; Viehrig, et al., 2012). However, the components of the dimension "task complexity" are based on the work by Ben-Zvi Assaraf and Orion just like the components of dimension 1 of the HEIGIS model.

Not all of the fields are covered by the studies. For instance, there is only one content and regional area and thus also little variance with regard to context familiarity. Moreover, after an initial survey of the data it was decided to drop level four of task complexity. The area of attitudes/motivations/interest is not covered in all studies, especially not in study 5. Based on the first study, spatial thinking was included as an important variable.

5.2 Item basis

Inspiration of how items can be formulated can be drawn from numerous previous studies. This section will give a general overview because there are many similarities between different studies, e.g. "X is interesting" in various grammatical variations to elicit student interest. The next section will include the details on the items and – in some cases – references to specific item sources. The coding presented here partly differs from earlier presentations.

For the affective domain, a general idea of how such items could be formulated, such as finding X interesting or occupying oneself with X outside of class, can be gained from numerous studies (e.g. Baker, 2002; Ballantyne, 1999; Bayrhuber, et al., 2002; Bednarz & Bednarz, 2008; Hemmer, et al., 2007; Hemmer & Hemmer, 1997, 1998; Hemmer & Hemmer, 2002a; Hemmer & Hemmer, 2002b; Hemmer, 2000; Hildebrandt, 2006; Kersting, 2002; Klein, 2007; Kunter, et al., 2002; Lee, Johanson, & Tsai, 2007; Lichtenstein, et al., 2008; Meyer, 2003; Obermaier, 1997, 2002b, 2005; OECD, 2000, 2003a, 2005; Orion & Cohen, 2007; Pieter, 2003; Reinfried, 2006; Rheinberg, Vollmeyer, & Burns, 2001; Rheinberg & Wendland, 2003; Schleicher, 2002b; Sommer, 2005; Storie, 2000; Uphues, 2007; West, 2003). Moreover, the "X" was partly inspired by looking at the national educational standards (DGfG, 2007a), e.g. the differentiation between physical geography, human geography and human-environment interactions and possible examples for topics within them.

With regard to systemic thinking the items are based on a survey of existing systemic thinking and spatial analysis items and partly learning tasks (e.g. Aarons, 2003; Baker, 2002; Ben-Zvi Assaraf & Orion, 2005a, b; Frischknecht-Tobler, Nagel, et al., 2008; Hildebrandt, 2006; Hlawatsch, Hildebrandt, et al., 2005; Hlawatsch, Lücken, et al., 2005; Kali, et al., 2003; Kaminske, 1996; Ker-

ski, 2000; Lee, 2005; Maierhofer, 2001; Ossimitz, 1996, 2000; Ossimitz, et al., 2001; Raia, 2005; Rieß & Mischo, 2008a; Schecker, et al., 1999; Sell, et al., 2006; Sommer, 2005; Sweeney & Sterman, 2000). Moreover, inspiration on different aspects that might need to be included were drawn from the national educational standards (DGfG, 2007a).

Spatial thinking has been included as a variable after the results of study 1. Spatial thinking is central to geography education (e.g. Rhode-Jüchtern, 2009; Siegmund, et al., 2012; overview in Viehrig, et al., 2012). The items were based on a survey of existing spatial thinking articles, both from the geo-sciences and other domains (e.g. Battersby, Golledge, & Marsh, 2006; Bednarz & Bednarz, 2008; Black, 2005; Crews, 2008; Gersmehl & Gersmehl, 2006, 2007; Golledge, Marsh, & Battersby, 2008; Hannafin, Truxaw, Vermillon, & Liu, 2008; Hooey & Bailey, 2005; Huynh, 2008; Jo & Bednarz, n.d.; Kali & Orion, 1996; Kerski, 2000; Lee, 2005; Lee, n.d.; Leopold, Górska, & Sorby, 2001; Marsh, Golledge, & Battersby, 2007; Montelllo, et al., 1999; National Research Council, 2006; Newcombe, 2006; OECD, 2006a; Oldakowski, 2001; Olkun, 2003; Prenzel, et al., 2004; Quaiser-Pohl, et al., 2006; Quaiser-Pohl & Lehmann, 2002; Self & Golledge, 1994; Sharpe & Huynh, 2008; Strong & Smith, 2001; Titze, Heil, & Jansen, 2008; Yang & Andre, 2003). Furthermore, there is a wide array of works regarding, for instance, general spatial thinking skills, spatial learning styles, spatial problem solving, spatial intelligence and spatial reasoning (cp. e.g. Diezmann & Watters, 2000; Freksa & Röhring, 1993; Gardner, 2005; Kreger Silverman, 2000, 2005; Li, Arbarbanell, & Papafragou, 2005; Mann, 2005; National Research Council, 2006; Sword, n.d.).

Pre-experience/pre-knowledge and preference questions (which are sometimes straight-forward) and can be found in previous studies (e.g. Bednarz & Bednarz, 2008; Carneiro, 2006; Hemmer, 2000; Hildebrandt, 2006; Klein, 2007; Lensment, 2002; Maierhofer, 2001; Neumann-Mayer, 2005; Obermaier, 2005; OECD, 2005; Orion & Cohen, 2007; Schleicher, 2002b; Sommer, 2005).

5.3 Study 1

For study 1, items regarding the pre-experience with Kenya and GIS; attitudes/motivations/interest in Kenya, computers, GIS and other media as well as

geography/geographic systems; computer skills; and several systemic thinking tasks are included.

5.3.1 Systemic thinking

Before the systemic thinking items, there is a short introduction to tourism in Kenya and a fictitious situation about a school newspaper wanting to report about the consequences of a hotel being build in a Kenyan village, including a map (based on GIS data). The context "school newspaper" is used throughout, and is somewhat similar to those used in other studies, such as the exhibition of a student work group (used in Hlawatsch, Lücken, et al., 2005), a researcher/consultant context (used in materials by Sell, et al., 2006) or a city committee (used in the factory inventory of Ben-Zvi Assaraf & Orion, 2005a). The 'should .... be build' question is also similar to the factory inventory of Ben-Zvi Assaraf and Orion as well as the learning tasks developed by Learning Africa (n.d.) and Clough & Holden (2002).

After the introduction, the first task is a multiple-choice item with five choices of which three are correct (relevant elements). In the study by Ben-Zvi Assaraf and Orion (2005a) students had to ask a list of experts questions. In general, multiple-choice questions are included in a number of systemic thinking studies (e.g. Hildebrandt, 2006; Ossimitz, 2000; Rieß & Mischo, 2008a).

The second is a relationship completion task, similar to the one used by Sommer (2005). Ben-Zvi Assaraf and Orion (2005a) also use pairs of relationships in preparation for constructing a concept map.

The third task (a) is a task asking for the evaluation of several spatial distributions (created with the GIS data) in terms of their similarity, similar to the item used by Lee (2005). Students also have to make a statement regarding the connections they assume between the displayed topics (3b). For instance, Baker (2002) has a task asking students first to address similarity and then explain why.

In the fourth task, students have to evaluate the transferability of conclusions from one location (Kenya) to another (Mali). The cover of the test features, besides photos, a satellite image with country boundaries (ESRI base map) where, apart from Germany, both countries (Kenya and Mali) are marked. Re-

motely similar tasks (e.g. consequences of whether it had rained or not, living in a room or outside) are used by Sommer (2005). In the book by Frischknecht-Tobler et al, which includes a review of task formats used in systemic thinking studies as well as several new empirical studies, tasks asking to compare systems are also included (2008).

The fifth task is to draw a concept map of 'tourism in Kenya'. A sample map of another topic (school) was given, similar to the procedure in other studies and materials such as Sommer (2005) or Hlawatsch et al. (2005). No concepts or relationships are given. Different forms (e.g. free drawing, with the help of given elements etc.) of concept maps and similar graphics (e.g. cause-/effect diagrams with only +/- as descriptions of the relationship) are frequently used in systemic thinking studies (e.g. Ben-Zvi Assaraf & Orion, 2005a; Frischknecht-Tobler, Nagel, et al., 2008; Hlawatsch, Hildebrandt, et al., 2005; Kaminske, 1996; Maierhofer, 2001; Ossimitz, 2000; Rieß & Mischo, 2008a; Schecker, et al., 1999; Sommer, 2005). Besides systemic thinking studies, concept maps are also used e.g. to study pre-concepts and conceptual change (e.g. Liu, 2004; Rebich & Gautier, 2005; Wallace & Mintzes, 1990) as well as students' knowledge (e.g. Fiene, 2014; Stracke, 2004) (for overviews of concept mapping see e.g. Åhlberg & Ahoranta, 2002; Fiene, 2014; Ruiz-Primo, 2004; Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005).

The sixth task is an evaluation of consequences, similar to the school change question in Sommer (2005) but asking not only for a list but also for reasons. 'What will happen ...'-type questions are also included in other studies (such as Hlawatsch, Lücken, et al., 2005; Maierhofer, 2001; Ossimitz, 2000).

The tasks are coded binary. Originally, level 1 was meant to be covered by item 1, 3, 5 and 6; level 2 by item 2, 3, 5 and 6, level 3 by item 4, 5 and 6 and level 4 by item 5 and 6. Expert feedback suggested that using one task for more than one level has to be seen very critically since item then are not independent and thus was abandoned in further studies. Only one level per task will be reported here, i.e. level 1 by item 1 and 3a, level 2 by item 2 and 3b and level 3 by item 4, 5 and 6. Level 4 is excluded from this reporting because what would constitute "hidden elements" for level 4 is, in the case of Kenya, not as clear as in the case

of other topics such as the water cycle. A survey of the students' answers shows that in general, very few include something that could be classified as 'not on the surface' (e.g. the contribution of research or education) in their answers. Due to this and the poor scale values of the other items which point to the need for more items, it was decided to exclude this level from further studies.

All items together have a poor value for Cronbach's alpha with 0.56 (0.56 if task 3a is excluded). Forming scales of items of one level to see whether they are similar shows that the level 1 scale has an unacceptable value with 0.20, suggesting that task 3a measures something different. The level 2 and 3 scales have poor values with 0.60 and 0.51 respectively. It has to be taken into account that the dichotomous coding of the answers could be a possible limitation (SPSSX Discussion, 2009). Shapiro-Wilk-tests for one sample showed that neither the average score of all items (p=0.01; M=0.40, SD=0.20), nor the average score without item 3a (p=0.004; M=0.40, SD=0.20) are normally distributed. The means of the individual items range from 0.10 to 0.91.

To further explore the achievement in systemic thinking and the item properties, a dichotomous Rasch model was applied to the items (see appendix). The final deviance of the model (all items) is 1104.1 (estimated parameters=11). The items have a weighted MNSQ between 0.93 and 1.09 and t-values between -0.8 and 1.1 and thus are within the acceptable range. The unweighted MNSQ values are between 0.80 and 1.13 (t -1.5 and 0.9). The EAP/PV reliability is 0.57, the WLE person separation reliability 0.52 and the coefficient alpha 0.56. Sufficient discrimination is achieved in all items.

Another Rasch analysis without item 3a (final deviance=965.6, estimated parameters=10) leads to weighted MNSQ values between 0.95 and 1.11 as well as t-values between -0.6 and 1.1. Unweighted MNSQ values are between 0.93 and 1.14 (t -0.4 to 1.0). The discrimination is acceptable. The EAP/PV reliability is 0.58, the WLE person separation reliability 0.52 and the coefficient alpha is 0.56. The item map can be seen in Fig. 16. It shows that the theoretically assumed levels can not be observed. However, the individual items of task 2 are increasingly more difficult, as assumed.
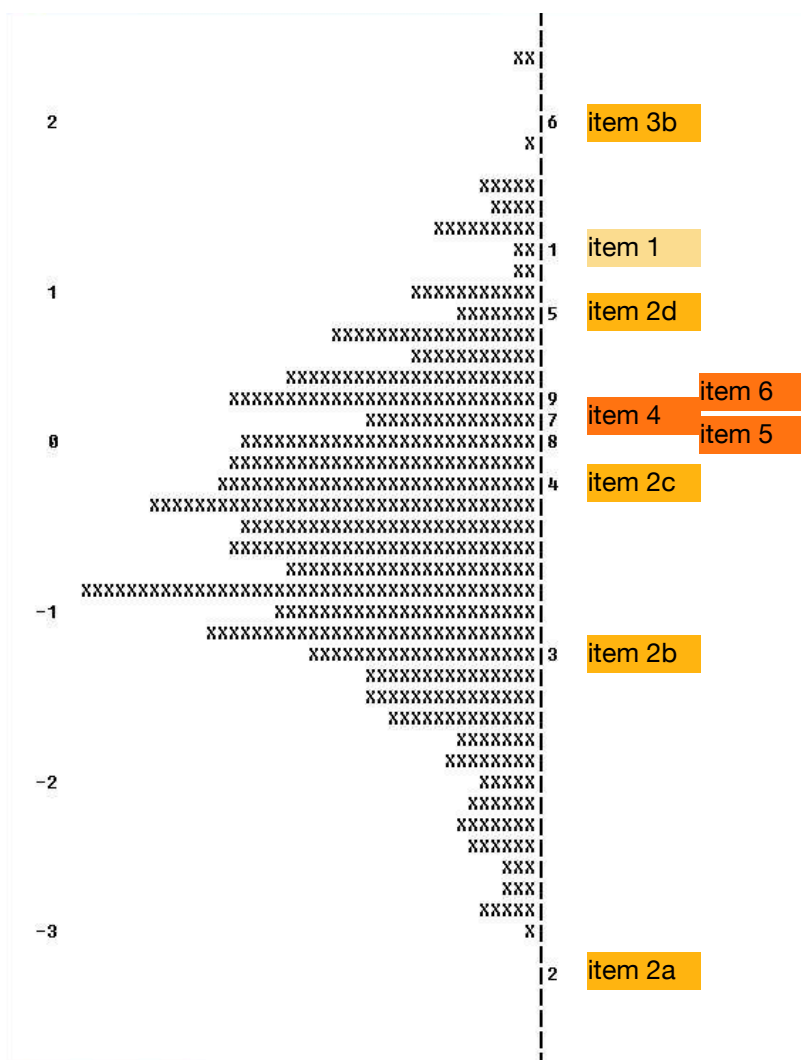
Fig. 16 : Item map: systemic thinking without item 3a
(light orange: level one, medium orange: level two, dark orange: level three)

For further analysis, the WLE estimates (without item 3a) are exported from Conquest and imported into SPSS. The WLE value are transformed with the help of the formulas WLEb=WLEa*(100/WLEa_SD) and WLEc=WLEb+(500--Mean_WLEb) from their original M=-0.56 and SD=1.33 to a M=500 and a SD=100, similar to the one used in the PISA study. The resulting value is not normally distributed (Shapiro-Wilk $p<0.001$) for the whole sample.

5.3.2 Conclusions

With regard to item development for geographic system competency assessment the study shows that coding open-ended tasks is a major issue. Moreover, item 3a (spatial/map-based identification of elements) might measure something different than the non-spatial element identification task 1. This is an

issue that needs to be further explored. Furthermore, the resulting scale (without item 3a) shows good Rasch scale properties, but only has a poor value for Cronbach's alpha, like many achievement tests. Additionally, the theoretically assumed levels can not be empirically verified. Due to the small sample size and the item issues, this might, however, be a product of test design.

5.4 Study 2

For study 2, items regarding attitudes/interest/motivations towards geography/ geographic systems, geo-media, GIS and Kenya as well as questions regarding the GIS pre-experience and several spatial thinking tasks are included.

5.4.1 Spatial thinking

Due to the results of study 1 some spatial thinking items are included. Tasks 77, 79 are adapted from Lee (n.d.), task 78 from Kerski (2000), and task 80 from Lee (2005).

The tasks seem to be fairly difficult for the students (means range between 0.17 and 0.60). The teacher feedback also hints at the tasks being considered fairly difficult, especially with regard to their graphics. The Cronbach's alpha of all four items is unacceptable with 0.35, it rises to 0.38 when item 77 is excluded and 0.54 when item 76 is excluded as well. The average score of all four items (M=0.45, SD=0.27, Mdn=0.50, n=98) is not normally distributed (Shapiro-Wilk p<0.001).

To further explore the achievement and the item properties, a dichotomous Rasch model was applied to the items. The final deviance of the model (all items) is 484.85 (estimated parameters=5). The items have weighted MNSQ values between 0.97 and 1.04 and t-values between -0.3 and 0.3, as well as unweighted MNSQ values between 0.94 and 1.18 (t-values -0.4 to 1.3) and thus are within the acceptable range. The EAP/PV reliability is 0.36, the WLE person separation reliability is 0.03, coefficient alpha is 0.35. The lowest item discrimination is 0.42. The item map (Fig. 17) shows that the items differ considerably in difficulty, especially the first two.
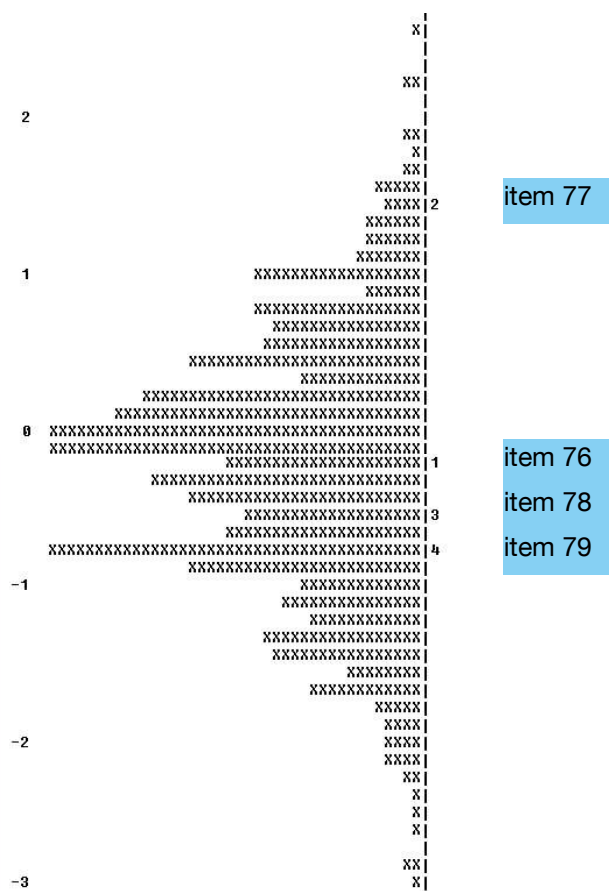
Fig. 17 : Item map: spatial thinking

For further analysis, the WLE estimates are exported from Conquest and imported into SPSS. The WLE value are transformed with the help of the formulas WLEb=WLEa*(100/WLEa_SD) and WLEc=WLEb+(500-Mean_WLEb) from their original M=-0.26 and SD=1.31 to a M=500 and a SD=100. The resulting value is not normally distributed (Shapiro-Wilk $p<0.001$) for the whole sample.

5.4.2 Conclusions

The spatial thinking items show, while fitting the Rasch model, very low values for Cronbach's alpha and are not easy for the students. Teacher feedback hinted at the graphics being a possible reason.

5.5 Study 3

The study includes the variables: spatial thinking, systemic thinking and the affective domain.

5.5.1 Spatial thinking

The tasks of study 2 are used again but in a different order and with a partly different wording in an effort to simplify and improve them. Additionally, the explanation is left out, as well as one of the choices in the US tasks (making both a choice of one out of four maps). Considering all four items together in the pretest, Cronbach's alpha is negative (-0.09) in the pretest (n=63). When item 4 is excluded, Cronbach's alpha is unacceptable with 0.41, this rises to 0.58 (poor) when item 3 is excluded as well. The average score of all four items is not normally distributed (Shapiro-Wilk, p<0.001; M=0.63, SD=0.17). Means range from 0.14 to 0.94.

Part of the difficulty of the youth club item (item 4) might be explained by problems with map interpretation. For instance, 23 students (36.5%) chose location "B" which lies in a commercial area bordering a residential area – the high failure rate might therefore be partly caused by the students being unable to transfer the location of the boundary correctly.

To further explore the items, a dichotomous Rasch model was applied to them. This results in a final deviance of 203.5 (number of estimated parameters=5). The coefficient alpha is -0.09, EAP/PV reliability is 0.02 and WLE person separation reliability 0.00. While all four items have good fit values (unweighted MNSQ between 0.97 and 1.04, t -0.1 to 0.3; weighted MNSQ 1.01, t 0.1 to 0.2) and thus can be assumed to be unidimensional, the item discrimination of the youth club item (item 4) is too low (0.14).

If item 4 is excluded, the final deviance is 144.3 (number of estimated parameters=4). EAP/PV reliability is 0.46, coefficient alpha 0.41 and WLE person separation ability 0.00. However, error messages indicating convergence trouble occur when exporting the WLE values. Unweighted fit values are between 0.62 and 1.01 (t -2.5 to 0.1), weighted fit between 0.91 and 1.08 (t -0.2 to 0.5).

A further reduction in items is not feasible. Due to the unfavorable statistical properties and the very small number of items the value of further analyses conducted with the WLE estimates calculated seem very questionable.

5.5.2 Systemic thinking

Compared to study 1 there is a greater differentiation between the three areas "human geography", "physical geography" and "human-environment interaction" (DGfG, 2007a). After a short introduction and an overview map of Kenya (based on GIS data), there are three multiple choice items for level 1, where students had to identify the maps that might help them to find more about a specific topic, similar to study 1. The level 1 items are coded so that a student needed to identify all choices deemed correct and none of those deemed wrong in order to get full credit.

Then there are eight statements such as "An increase in tourism has no effects on the local culture" where students have the option to check one of four options ranging from "don't agree at all", "rather don't agree", "rather agree" to "totally agree". Agreement/disagreement items were used e.g. by Ben-Zvi Assaraf and Orion (2005a, b). The level 2 items are first re-coded where necessary in case of reverse items, and then classified as "2" and "3" (rather agreeing) given credit and "1" and "0" (rather not agreeing) or no answer given no credit. In the same block, there are two additional items (i16 and i17) that have a time component, as another attempt to shed light on level 4, although they could be also seen as a variant of level 2 (relationship of a variable with itself which includes a threshold). 'What would happen' tasks with possible thresholds have been used e.g. by Maierhofer (2001) and Ossimitz (1996). They are coded the same way as the other statement items.

Then follow three level 3 tasks, which consist of concept maps. In contrast to study 1 there is a pre-existing concept map in which three given terms need to be connected to another term, without providing the students with given relationships. In order to get credit, the connection needs to have a reasonable label and correct arrowhead. A connection to only one term is needed, as no minimum number of arrows is specified. This leads to the items being a borderline case between level 2 and 3, strictly speaking, since it then only requires an individual relationship, although it is embedded into a more complex network of relationships. This does not change if every map is considered as an item

instead of every term (labelled "all task"), albeit it would then be more difficult to get credit (needing all three connections instead of just one).

Due to the empirical differences in the two similar items (item 16: M=0.67, SD=0.48, item 17: M=0.17, SD=0.38) and the conceptual difficulties, items 16 and 17 are excluded from further analyses.

Without item 16 and 17, the scale has a good value with 0.87, dropping to a questionable value with 0.62 if each map is considered as an item instead of each term. Step-by-step exclusion of items till no further improvement possibility is indicated leads to a scale of only level 3 items (18 a and b, 19 a to c, 20 a to c, alpha 0.94) for the term option and a scale of level 2 and 3 items (8, 10, 18task, 19task, 20task, alpha 0.77) for the task option. That means a lower value of Cronbach's alpha must be accepted in order to cover all three levels. Consequently, for the raw score analysis, all items (with task is item for level 3) except item 16 and 17 are used (pretest M=0.60, SD=0.17, posttest M=0.65, SD=0.19, difference M=0.05, SD=0.12). Shapiro-Wilk tests show that only the posttest is normally distributed (raw pretest p=0.021, posttest p=0.062, difference score p=0.014).

It is notable that the items do not display model conform frequencies, with level 2 items being the easiest (M between 0.75 and 0.92), level 3 being in the middle (task: M between 0.25 and 0.41, term 0.35 to 0.62) and one of the level 1 items being the hardest (M between 0.10 and 0.54). This might partially be explained by the used item formats. The unexpected difficulty of the level 1 items might be immanent in the tasks themselves, since students have to evaluate five elements correctly in order to get full credit. An individual coding of each element is not possible, however, because the options were only 'checked' or 'unchecked', not 'helpful' or 'not helpful', i.e. students who have not worked on the task would get credit for a correctly 'unchecked' item. It is also notable that the third term in a concept map is slightly more difficult than the others, as expected (item 18: M_a=0.62, M_b=0.44, M_c=0.35, item 19: M_a=0.60, M_b=0.60, M_c=0.48, item 20: M_a=0.56, M_b=0.60, M_c=0.46). Furthermore, level 3 items are more difficult than level 2 items. On the whole, it seems that the physical geography map (task 18:

M=0.25) is more difficult than the human geography (task 19: M=0.41) and human-environment interaction (task 20: M=40) maps.

To further explore the items, a simple Rasch model was applied to them. Using each term as an item, it results in a final deviance of 1240.1 (number of estimated parameters=21). The coefficient alpha is 0.87, EAP/PV reliability is 0.85 and WLE person separation reliability 0.82. Five items have an item discrimination smaller than 0.25. The fit values of many of the items are not good. While 1 of the level 1 and 4 of the level 2 items have weighted MNSQ values that are too high, 7 of the level 3 items have values that are too low. Eight items have t-values out of range. With regard to unweighted fit, two of the level 1 items and five of the level 2 items have MSNQ values that are too high, and all level 3 items values that are too low. T-values are out of range for 14 items. These results might point to a possible divide between 'passive' tasks that require students to make a decision regarding a given matter and 'active' tasks that require students to write answers themselves. However, the sample size seems inadequate for multidimensional modeling and the tasks too confounded, since active tasks were only required on level 3.

Another option would be to use the 'task as item' format. The final deviance is 899.2 (number of estimated parameters=15). The EAP/PV reliability is 0.63, the WLE person separability 0.61 and the coefficient alpha 0.62. The weighted MNSQ values are within the range for all items, however, two items have an out-of-range t-value. Four items have an item discrimination below 0.25, three items have unweighted MNSQ values outside of the range, with one t-value being too high. To arrive at fitting items, several possibilities exist. Thereby, the task option seems to be a better starting point than the term option.

First, item 13 is excluded, because it has the lowest discrimination (0.02) as well as too high unweighted MNSQ/t-values. Also excluded are items 19 (too low unweighted MNSQ, too high weighted t-values), 5 (too high weighted t-value, low discrimination), 14 (too high unweighted MNSQ, low discrimination) and 15 (low discrimination). There remain nine items (items 6, 7, 8, 9, 10, 11, 12, 18t, 20t).
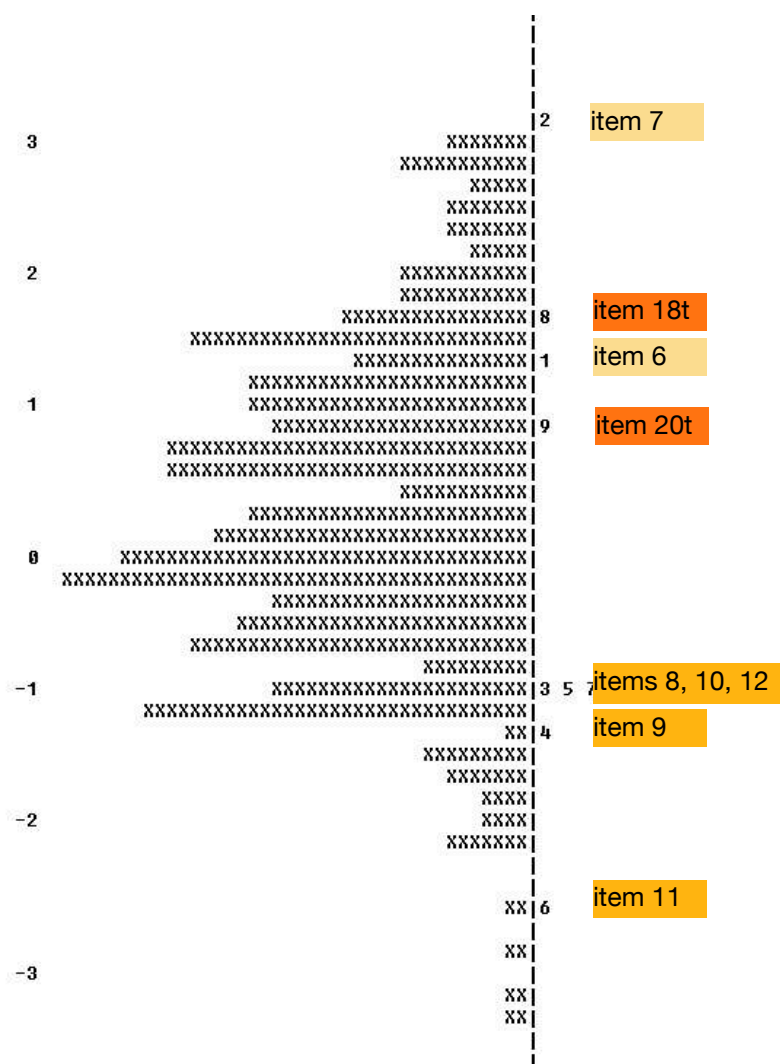
Fig. 18: Item map: pretest systemic thinking with items excluded
        (level 1: light orange, level 2: medium orange, level 3: dark orange)

The analysis is run again with the remaining items (see Fig. 18) and exporting the necessary parameters, which are imported to run the posttest analysis (see appendix). In the pretest, the final deviance is 555.1 (number of estimated parameters=10). The coefficient alpha is 0.62, the WLE person separation reliability 0.60 and the EAP/PV reliability 0.64. The final deviance of the posttest is 551.1 (estimated parameters=0). The WLE person separability is 0.54, the EAP/PV reliability 0.58 and the coefficient alpha 0.59.

Both pre- and posttest WLE estimates are imported into SPSS to be used for analysis. The mean and standard deviation of the pretest scores are transformed with WLEb=WLEa*(100/WLEa_SD) and WLEc=WLEb+(500-Mean_WLEb) from their original M=0.33 and SD=1.49 to a M=500 and a SD=100. The posttest scores are transformed with the same values from their original

M=0.83 and *SD*=1.42 to a *M*=533.59 and *SD*=95.61. The difference score has a *M*=33.59 and *SD*=73.95.

Shapiro-Wilk tests show that all but the pre-score are normally distributed (WLE pretest *p*=0.030, posttest *p*=0.054, difference score *p*=0.073).

5.5.3 Conclusions

Study 3 shows that there is still considerable need for development of the spatial thinking scale. With regard to the systemic thinking tasks, the study confirms the decision to concentrate on the first three levels. The study results are very limited due to the small sample size but hint at difficulties to get a good unidimensional scale. The study also highlights the results of differences in terms of data coding and item inclusion criteria. This underscores the need for a broad range of extensively empirically validated measurement instruments for different levels and topics in order to be able to explore the effects of using different media and methods.

5.6 Study 4

The study includes only spatial thinking items.

5.6.1 Spatial thinking

The spatial thinking items contain the same four items as study 3 (item 1, item 2, item 3 respectively 4, item 4 respectively 5). Additionally, the second version has a number of simplified items. These are a simplified and location unspecific generalization of the USA item (correlation: simplified, item 3_s), two map overlay items (similar to GIS layer principle explanations, see e.g. ch. 1; Siegmund, 2001), where students have to specify the correct solution if two or three, respectively, transparent foils are laid on top of each other (item 6a_s, item 6b_s), three farmer's task items where students have to specify, based on two schemas (one which shows the areas belonging to farmer A and farmer B and another what is grown on the areas, items 7a_s, 7b_s, 7c_s), the correct of three colored areas based on a description such as "Which part of the area is planted with wheat and belongs to farmer A?", requiring both map overlay and basic logic operator skills (similar to

Battersby, et al., 2006) and a simplified version of the youth club task where students have to identify the correct of 5 locations based on 3 conditions (item 8_s).

Compared to study 2 and 3, the four tasks with the more complex graphics seem to be easier for the sample, however, the youth club task is still difficult (Fig. 19). Considering the sample, however, which consists predominantly of geography students, the results are not that good. The simplified items are, indeed, much easier. A Wilcoxon test for dependent samples shows the difference between the average of all simplified items and the four complex ones is significant ($p<0.001$). Some students' drawings hint at how they solved it, and point to, despite the graphics/printing, that the difficulties might not only be in the thinking skill necessary to solve a task, but partly also in reading the text, legend and graphic.

Shapiro-Wilk tests show that the raw average score of all items ($M=0.82$, $SD=0.12$) is not normally distributed ($p<0.001$, $n=145$). Neither are the raw average scores of the four old ($M=0.69$, $SD=0.22$, $p<0.001$, $n=190$) or all simplified items ($M=0.89$, $SD=0.13$, $p<0.001$, $n=145$). The Cronbach's alpha of all items is unacceptable (0.34), as is that of only the four "old" items (0.31) and all simplified items (0.31). A ceiling effect could be part of the reason for this. Step-wise exclusion of items leads to a Cronbach's Alpha of 0.41 for seven items, including two old (item 2 and 3/4) and five simplified ones (item 3_s, 6a_s, 7a_s, 7b_s and 8_s).

To further explore the items, a simple Rasch model was applied to them ($n=145$). The final deviance is 1196.62 (number of estimated parameters=12). The items are within an acceptable range and thus can be assumed to be unidimensional (unweighted MNSQ 0.84 to 1.13, t -1.4 to 1.1, weighted MNSQ 0.96 to 1.04, t -0.4 to 0.3). The coefficient alpha is 0.34, WLE person separation reliability 0.10 and EAP/PV reliability 0.32. All items have an acceptable item discrimination, the lowest being 0.26. The item map (Fig. 19) shows the ceiling effect. The WLE estimates are imported into SPSS and transformed with WLEb=WLEa*(100/WLEa_SD) and WLEc=WLEb+(500-Mean_WLEb) from its original $M=1.87$ and $SD=1.10$ to a $M=500$ and a $SD=100$. A Shapiro-Wilk test shows that the WLE average score is not normally distributed ($p<0.001$).
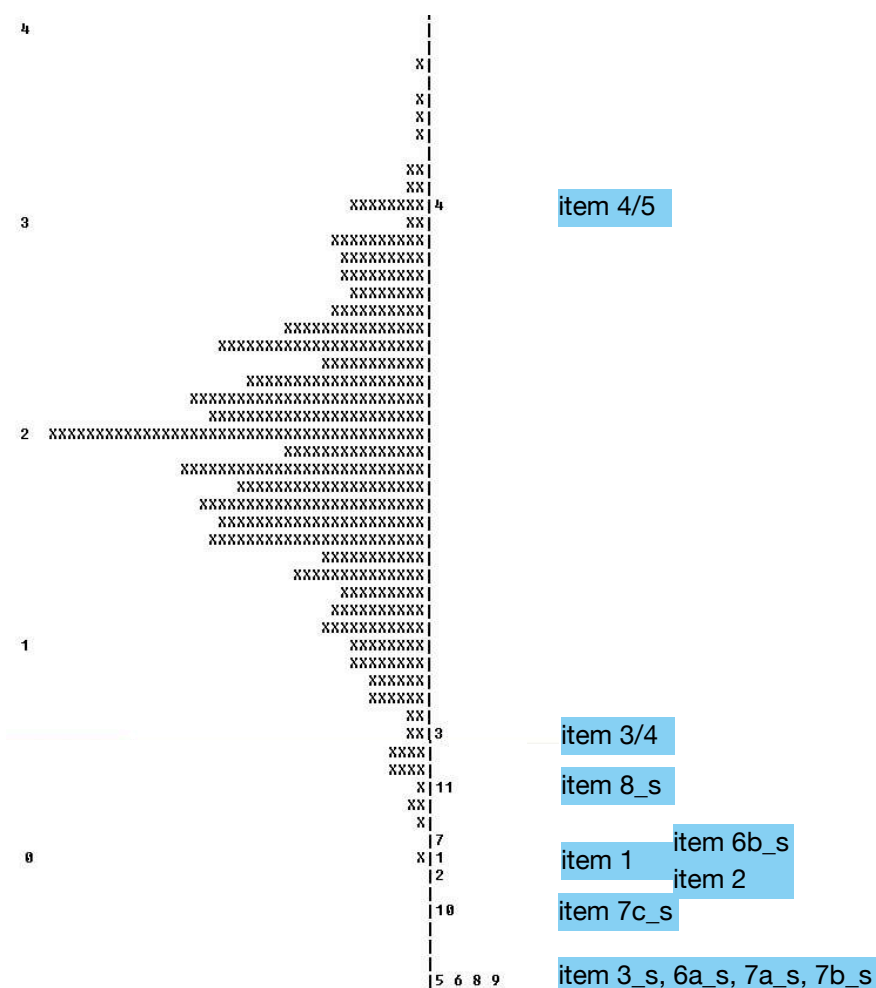
Fig. 19: Item map: spatial thinking

5.6.2 Conclusions

The study shows that spatial thinking items utilizing simplified graphics are indeed significantly easier than those using more complex ones. The simplification achieved actually is too great for the sample, producing a ceiling effect and accompanying limitations on the information value of further analysis. However, since this sample consists of university students, the simplified items could work with school students.

5.7 Study 5

In study 5, due to test time considerations, only few variables could be included:

- in the pretest: stream, sex, grade, age, family languages, other languages, birth country of the student, birth countries of the students' parents, spatial thinking, systemic thinking

- in the posttest: medium worked with and feedback, spatial thinking, systemic thinking

Additionally, in the materials, there are a few questions included that could be treated as variables, such as whether the students had already worked with GIS/atlas and if yes with which and what they already know about Kenya.

5.7.1 Spatial thinking

In contrast to studies 2, 3 and 4 only simplified items are used. Several items are taken from study 4, changed from the pronoun *Sie* generally used for adult students to the pronoun *Du* used for children and adolescents. Item 3 is used as item 17. Task 6 (the two transparency items) are used as task 13. Task 7 (the three farmer items) are used as task 18. Item 8 (the simplified youth club item) is used as item 20. Additionally, very simplified versions of the ice cream parlor item used in study 4 (item 4) (item 19 in study 5) and the correlation item (item 1 in study 4) (item 16 in study 5) are created. There are also a new even more simplified correlation item (only 2 choices, item 15) as well as three items dealing with correlation each with two choices (e.g. where there is a restaurant of of chain A there is also one of chain B, task 14). There are also two items where students have to evaluate the distribution of visitors in a park (task 12). Additionally, there are two items that deal with whether a point, line or polygon is the best representation of an object on a map of Germany (such as a weather station) (task 11) (Lee, 2005, also has a item that asks for spatial information display as point etc.).

Except for item 17, the items are relatively easy for the sample, as could be expected of simple tasks. However, in light of the very basic nature of some of the tasks it is surprising that there is still a considerable percentage that does not get a point (e.g. 6% in case of task 14a). Cronbach's alpha of all items is poor (0.58), if items 17, 13b are excluded, the value is 0.60 (the same as when 13a is excluded as well). A Shapiro-Wilk test shows that the average score with these three items excluded is not normally distributed ($p<0.001$).
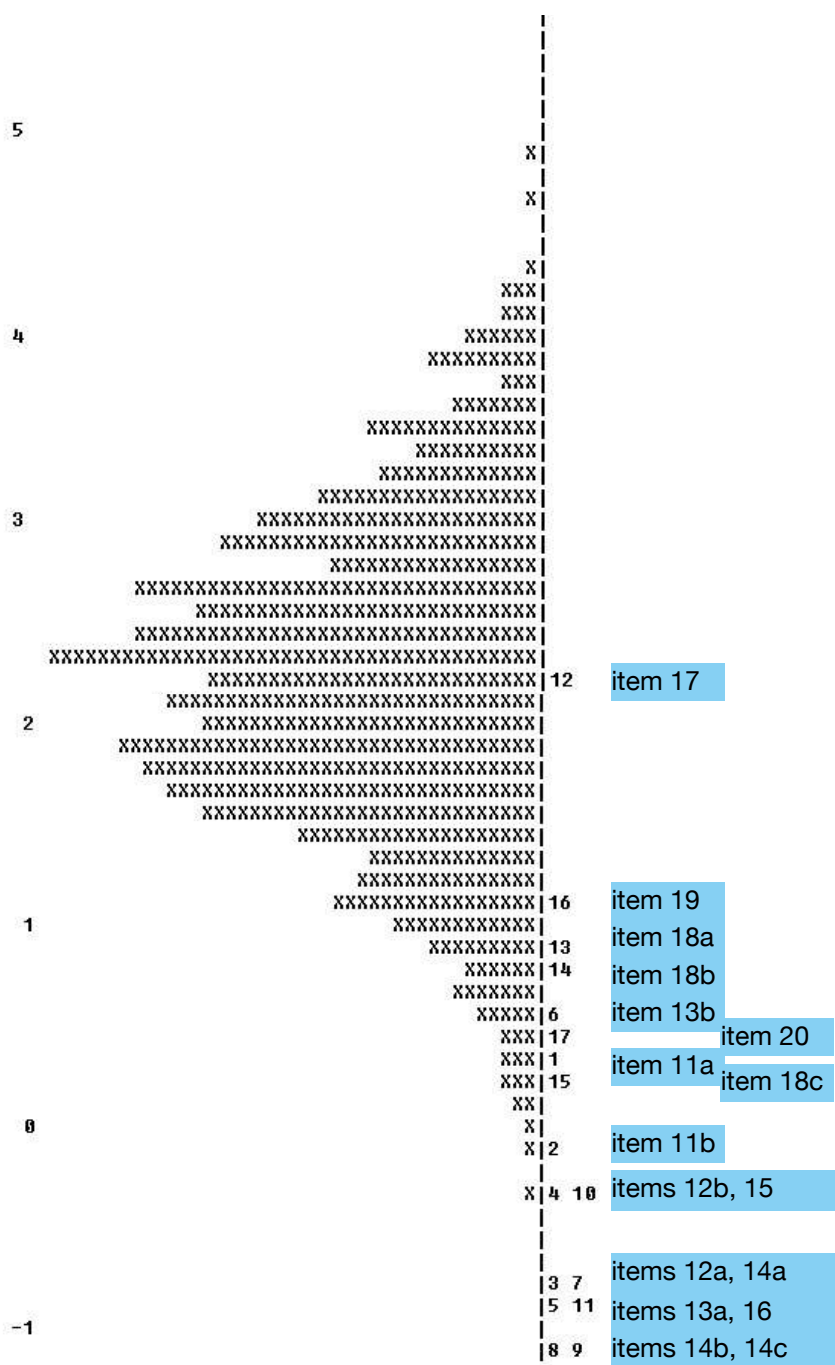
Fig. 20: Item map: spatial thinking

To further explore the items, a simple Rasch model was applied to them. The final deviance is 11563.8 (number of estimated parameters=18). The WLE person separation reliability is 0.28, the EAP/PV reliability 0.54 and the coefficient alpha 0.58. Overall, the items show acceptable fit and discrimination values, but partly high t-values (weighted MNSQ 0.96 to 1.07, t -1.0 to 3.1, unweighted MNSQ 0.84 to 1.17, t -3.6 to 3.5, discrimination above 0.25). However, the sample size here is considerable larger than the about 300 for which the parameters have been specified. Fit indices take sample size into account, the larger the sample

size, the closer the expected MNSQ value is to 1 and the more t-values are likely to be > |2| (Wu & Adams, 2007). Wu and Adams recommend that "[t]he reliability of the test and item discrimination indices should also be considered [...]" (p. 85). If all items having out of range t-values would be excluded (11b, 12a, 14a-c, 17), two items show an out of range t-value (13a, 13b), the coefficient alpha decreases to 0.55, the WLE person separation reliability to 0.03 and the EAP/PV reliability to 0.50. Thus, it seems best to use all items. The item map (Fig. 20) underscores how easy the items were for the sample.

The WLE estimates are imported into SPSS. The pre-WLE value is transformed with the help of the formulas WLEb=WLEa*(100/WLEa_SD) and WLEc=WLEb+ (500-Mean_WLEb) from its original M=2.14 and SD=1.05 to a M=500 and a SD=100. According to a Shapiro-Wilk test the score is not normally distributed (p<0.001).

5.7.2 Systemic thinking

After a map showing the location of Kenya in Africa and its neighboring countries (based on GIS data), the test contains eight items.

The first one (i22) is again a graphic based tasks, where students have to identify the climate zones relevant for Kenya and could use a map to help them (using the climate classification of Siegmund & Frankenberg, see e.g. Siegmund, 2008).

The second and third items (i23, i24) are multiple choice format, students having to check the one map out of four that could not help them. I23 is from the area of physical geography, i24 of human geography. Only answers that have a check mark for the correct option and none for any other options are given full credit.

There are three level 2 items (i25 to i27) where students have to read a statement, check whether it is correct or not, and explain their choice. This is a compromise between only checking their agreement/disagreement and completely open questions in the form of 'statement ... is it true' (see e.g. also Ben-Zvi Assaraf & Orion, 2005b; Frischknecht-Tobler, Nagel, et al., 2008). Statements with explanation are used e.g. by Ben-Zvi Assaraf and Orion (2005a) and Sommer (2005).

In order to get credit for the level 2 items, students have to check the correct response (i.e. "yes, that's correct" for items 25 and 26 and "no, that's not cor-

rect" for item 27) and – to reduce the 50% guessing chance – not leave the 'why' answer blank but give an explanation that indicates that he/she has not completely misunderstood or guessed the question. For example, item 25 is ""National and international political developments can influence the tourism in a location." - Do you think this statement about Kenya is true?". An answer getting no credit would include writing something like "no idea", but also for instance (all student examples translated):

- unclear answers such as "weather, coffee" (ID 0701, pretest)
- restatements such as "The politics has, I believe, also something to do with tourism" (ID 1114, pretest)
- general statements such as "I think Kenya is a relatively popular holiday destination" (ID 0944, pretest)
- answers contrary to their check mark such as "which tourists are interested in the politics of another country" (ID 0183, pretest)
- strange answers such as "Because they have another culture (fish, pencil case, cake)" (ID 0973, pretest)
- answers like "It depends on the place" (e.g. ID 0767 posttest) since there are multiple ways of influence that could be used as examples and it's not a real answer to the "why question"

and a number of individual answers not easily assigned a special type.

More complicated is the decision of what to with students giving only one or two word answers such as "war". These kind of answers do not provide an explicit link between a political development and tourism, but make sense (a more elaborate answer is: "because in case of a war e.g. in this country no tourists want to come" (ID 0202, pretest)). One word answers also occur in other areas such as "hotel" (a more elaborate answer is: "Because when the politics decide to build hotels, cinemas, swimming pools, etc. in a location in Kenya, the tourism increases" (ID 0468, pretest)). In between the two extremes (one word – explicitly described connection) there are many answers that are slightly incomplete, but can be understood e.g. "because they decree e.g. how many hotels are built" (ID 0963, pretest), which could be awarded a 1 when "they" is understood to mean politicians.

Overall, each of the different possible boundary lines has some cases that are doubtful and each possibility is introducing some kind of bias. For the coding, different strategies were tried out. On the whole, it needs to be stressed that the coding of these kind of tasks is very time intensive with the sample size achieved in this study. This also leads to a coding by multiple persons being not feasible. A small part of the sample was coded by a second person in a draft version of the coding, and the coding scheme discussed afterwards.

A very strict coding, i.e. only counting students that provide an explicit link between the areas of politics and tourism, would give a lot of students that seem to not have completely guessed a 0, against the original intention of the tasks (else one could have simply asked a completely open question such as "Are politics and tourism connected? Describe an example."). It gives a lot of weight to motivation (writing full sentences or at least phrases) and linguistic skills. This seems problematic, especially considering the troubles some of the students seemed to have with the German language, even if they specified that both they and both of their parents were born in Germany. This included not only grammatical mistakes and the wrong capitalization of words, but also misspelling some primary grade words such as "*lehrnen*" (correctly: *lernen*, translated: learn). Whether these were cases of students with dyslexia or simply without adequate development in German language skills cannot be determined from the test used.

The other extreme coding, i.e. even awarding a point for one-word-answers, such as "war", or incompletely specified phrases, such as the example with "they" above, is also potentially biased, because it only assumes that the students have understood the whole connection. Answers that were not clearly articulated or depending on interpretation could be right, but also could be plain wrong.

This problem cannot be resolved completely satisfactorily. Consequently, other item formats should be developed, especially for studies with a sample size above 100 and thus a large extent of variation in the clarity of arguments and themes (for item 25, the themes included e.g. security, media attention, laws, entry permits, attractions and finance as well as a general more/less tourists). For this report, it needs to be kept in mind, that the decision for one coding guideline has a big impact on the solving rates, and thus, potentially, the results. Because a choice

needed to be made, for this report, it has been decided to err in favor of the students, i.e. use a lenient coding.

For the final coding, student answers are first sorted into correct check mark – not correct check mark, the latter being awarded a 0. Then the answers of those with correct check mark but blank or "I don't know etc." are assigned a 0. Finally, the others are first categorized (e.g. "war", "restatement", "entry permits", "unclear") and then assigned a 0 or 1. This seems necessary since the amount of data collected was huge, leading to problems especially when – like in earlier coding – trying to assess the merit of a student answer one by one and only going back to similar cases when noticing. Classifying the student answer takes considerable time and is attached to a degree of uncertainty in some cases, but helps in managing the data amount to become more stringent in the coding, as only answers that are somewhat alike have to be compared in case of boundary cases.

Item 27, which was phrased negatively, was more difficult for the students than the two other items. In general, while negative phrasing is often used in research (e.g. Ben-Zvi Assaraf & Orion, 2005a; Hildebrandt, 2006; Moosbrugger & Kelava, 2007; Obermaier, 2005; Sommer, 2005; Uphues, 2007) it seemed very difficult for the students in this case.

Then, there are two level 3 items. In the first item students have to, after a short explanation what a concept map is, connect one term to two others in a prefabricated concept map through fittingly labelled arrows. In the second, students have to construct a concept map about the problems of tourism in Kenya with as many connections as possible from a list of eleven terms. No relationships were given.

For item 28, students have to connect the term "school" (not another term) to at least two other terms. The connections need to have an arrowhead in the correct direction. The pre-contemplation was:

- They need to be labelled in a way that was correct and 'readable'.

- Minor grammatical mistakes are allowed (e.g. need instead of needs).

- Relationships that are readable when read as continuation of another relationship are okay.

- Whole sentences are okay, irrespective of arrow direction, indicating correct conceptual understanding but lacking understanding of the format.

In general, what was considered still 'readable' and 'correct' enough was often a difficult decision. The coding needed to be stricter than for i25 to i27, because there the text only served to reduce the guessing probability. At the same time, the coding should not give too great a weight to language skills or format. Similar problems have been discussed by Yin et al. (2005, p. 181) who described among other things that

> "Since language skill was not our assessment target, we did not want students to lose credit because of their lack of language proficiency. Therefore, we had to make an informal judgment as to the intended meaning underneath their awkward wordings. As a result, numerous diverse linking phrases created by the students led to significant scoring challenges."

For the coding, consequently, it was decided that (all student answers translated):

- Answers that comprised not a verb, but a noun (e.g. "education"), a noun and an adjective etc. were counted as not sufficient, even though many of these could indicate correct understanding (e.g. "school ---education---> workers" ID 0366, pretest; "school ---good job---> money", ID 0012, pretest) and thus could be coded as 1 in the most lenient possible coding. Often, these could stem from an understanding of the arrow as "leads to". Exceptions of the insufficient coding were those that could be read in continuation, also in combination with the exceptions below.

- To not give an overweight to language, answers that could be made readable by adding prepositions or articles were allowed (e.g. "workers get job school", ID 0368, pretest, can be made readable to "workers get a job in school"). This results in also allowing answers referring to students or teachers since they could be phrased as "teachers of the school". This led to, while generally, the arrow direction was important, sometimes also having to accept a controversial arrow direction (e.g. "money needs the school" in German works as emphatic sentence, even though "money needs school" would be difficult). However, some arrows could not be reconciled. For instance, "workers ---education---> school" can be used, by combining the continuation and

preposition leniency (hotels need workers with education from school), while "school ---education---> workers" is not reconcilable.

- While in general, wrong connections were not counted, in the case of irreconcilable double arrows, they were counted as wrong.

- Another issue was missing parentheses (e.g. "school needs workers (teachers)" is doubtless correct, "school needs teachers workers" is not as clear). These were counted as correct. They could sometimes be alternatively resolved by prepositions, e.g. "school needs teachers as workers"

- The extent to which grammar mistakes were allowed was another issue. Moreover, spelling mistakes, when the word was still recognizable (e.g. "*bracht*" instead of *braucht*) were okay.

- Another problem were phrases in the middle between one word answers and complete answers, e.g. "school without education no money" (ID 0836, pretest) or "school to bring students (from out of town) to school infrastructure" (ID 0179, pretest). On the one hand, these students wrote more than the one word answers. On the other hand, the line between "phrase" and "not reconcilable few word answer" would be very fuzzy. Consequently, those were not counted as correct.

On the whole, these examples show the difficulty involved.

Moreover, one fairly common connection was that "school brings/creates workers" (sometimes also even less accurately "gives" or "provides"). While technically, people can work even without school, and frequently do so especially in developing countries, an experienced teacher noted that students get told all the time that they need a good school certificate in order to be able to work and thus this connection should be counted as correct even without a specification ("more qualified").

Another difficult connection occurring in different versions was whether tourists need to go to school or not. These were either connected correctly ("tourists visit schools") or would need to be true to count for continuation ("tourists need infrastructure (e.g. streets) to get to school", ID 0624 posttest). On the one hand, one student specified correctly that old or interesting schools attract

tourists. Moreover, some tourists may be interested in education in another country and thus try to visit schools, either through meeting teachers or as part of a village tour. On the other hand, schools in general are not 'normal' tourist destinations and thus these unqualified connections are difficult to count as correct. However, in favor of the students, these were counted as correct.

A further area of difficulty is "school brings tourists" or "tourists need school". While indirectly, one could think of possible connections (e.g. a better school education leads to better qualified people leads to a better situation in general and in particular better advertisement etc. to attract tourists leads to more tourists; or tourists need people to understand them, so they need people with education e.g. in English, ...). However, these were quite "roundabout" and intermediate steps were not spelled out, thus they were not awarded a point.

Furthermore, there were several other individual areas such as "tourists have kids, kids need school" (difficult because the concept map should be about tourism), which was scored based on the individual phrasing (e.g. some wrote that tourists move to Kenya), the school - vacation connection, etc.

Overall, there are three possible cut-off points:

(1) the pre-contemplation coding is the strictest coding (Pre: 353, Post 347), which also excludes "school brings workers"

(2) the compromise coding, which is not totally lenient but gives less weight to language as (1); only four (pre) or three (post) cases are worse than (1), 155 (pre) and 168 (post) better

(3) the lenient coding, which would also allow for nouns etc. where a possible connection could be construed etc.

For this report, coding (2) was chosen.

For item 29, students had to use at least six of the terms (which is slightly more than 50% of the eleven specified). Again, the students needed to use arrowheads in the right direction and a valid label. For the number of connections there were two choices:

(a) the same number as terms used, which would make both networks and cycles count; or

(b) one more connection than terms used, which would make only networks count

For this report, coding (a) was chosen. Students could use additional terms, however, these were counted for the number of connections required, e.g. a students using six of the terms and one additional term for his network would need seven connections to get full credit.
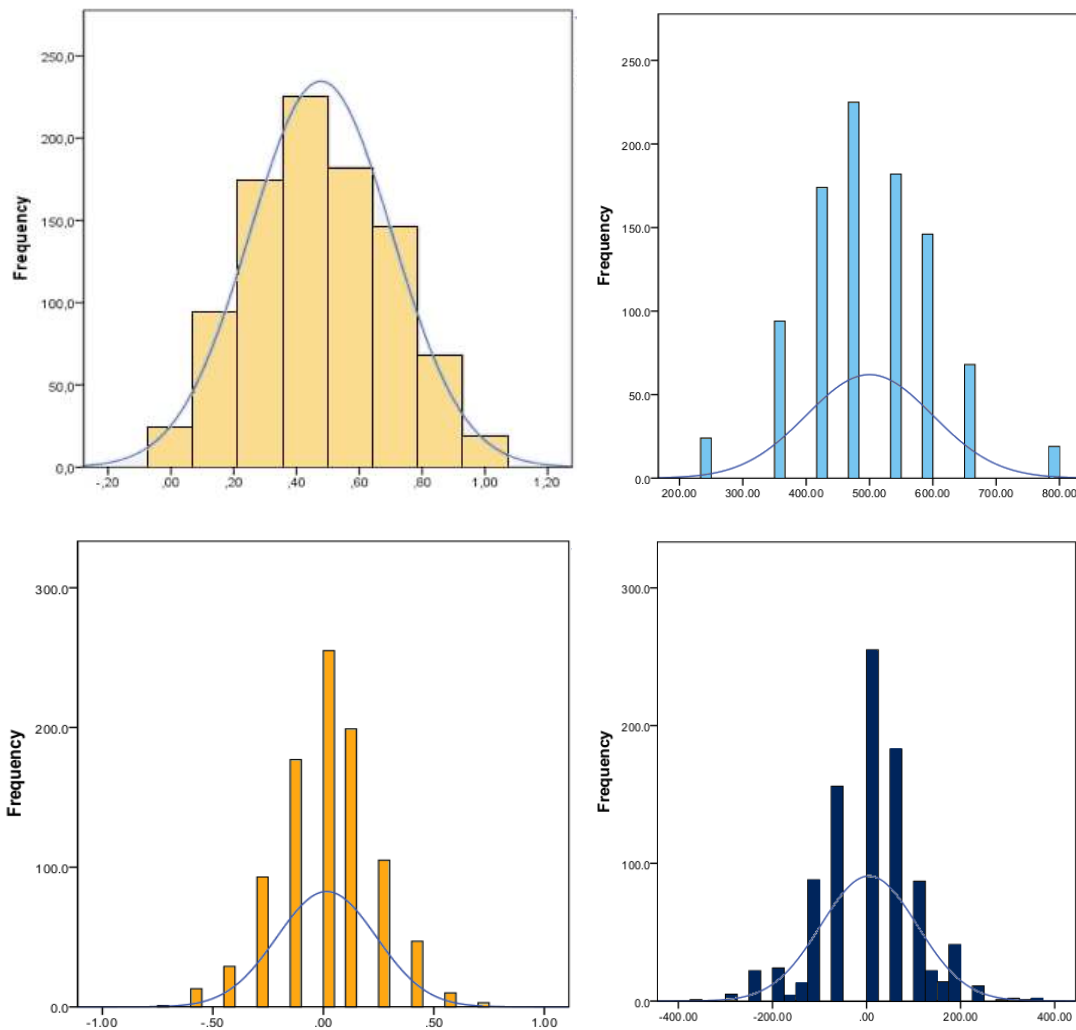
Fig. 21: Histogramms of the systemic thinking pretest and difference average scores
         (top left: raw pretest without item 22, top right: WLE pretest
         bottom left: raw difference without item 22, bottom right: WLE difference)

Cronbach's alpha of all eight items in the pretest is 0.42, if item 22 is excluded, the value is 0.48. The average score (M=0.47, SD=0.21) and the average score without item 22 (M=0.48, SD=0.23) are evaluated with Shapiro-Wilk tests, which show that both are not normally distributed ($p<0.001$). However, the graphical deviation does not seem that great (Fig. 21) (see also e.g. Kollar, 2012).

To further explore the items, a simple Rasch model is applied to them. The final deviance is 9098.7 (number of estimated parameters=9). The WLE person separation reliability is 0.33, the EAP/PV reliability 0.42 and the coefficient alpha 0.42. Overall, the items show acceptable fit and discrimination values (unweighted MNSQ 0.93 to 1.09, t -1.6 to 2.0; weighted fit 0.95 to 1.07, t -1.2 to 3.1; discrimination 0.28 to 0.52). Item 22 has a higher t-value (3.1 for the weighted, 2.0 for the unweighted fit) as well as a lower item discrimination than the other items (0.28).

If item 22 is excluded, the final deviance is 7797.4 (number of estimated parameters=8). The WLE person separation reliability increases to 0.36, the EAP/PV reliability to 0.48 and the coefficient alpha to 0.48. Overall, the items show acceptable fit and discrimination values (unweighted fit 0.93 to 1.05, t -1.6 to 1.0, weighted fit 0.97 to 1.03, t -0.9 to 1.1, discrimination 0.44 to 0.56).

The item map (Fig. 22) shows that the items are fairly well matched to the sample. However, item 28 was easier than expected. The item map also highlights the differences in difficulty between i25/26 on the one hand and i27 on the other hand.

Both pre- and posttest WLE estimates (without i22) are imported into SPSS. The mean and standard deviation of the pretest scores are transformed with $WLEb=WLEa*(100/WLEa\_SD)$ and $WLEc=WLEb+(500-Mean\_WLEb)$ from their original $M=-0.13$ and $SD=1.20$ to a $M=500$ and a $SD=100$. The posttest scores are transformed with the same values, from their original $M=-0.07$ and $SD=1.39$ to a $M=505.22$ and $SD=116.50$. Shapiro-Wilk tests show that the pre- and post-WLE scores are not normally distributed ($p<0.001$).

Difference scores (post-pre) are computed for both the raw ($M=0.02$, $SD=0.23$) and the WLE data ($M=5.22$, $SD=102.38$). Neither difference score is normally distributed according to Shapiro-Wilk tests ($p<0.001$).
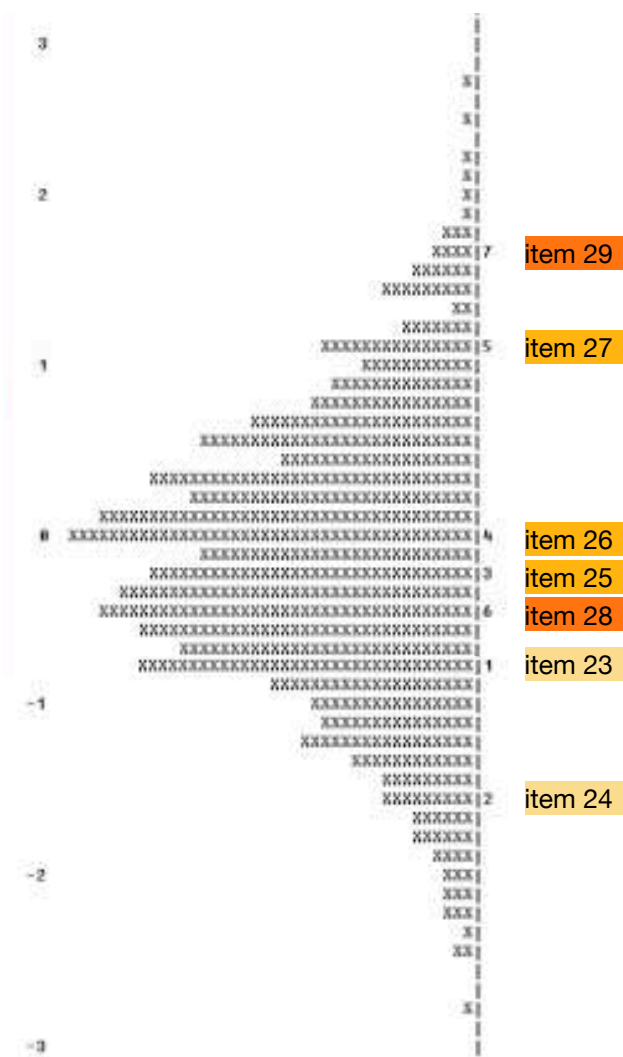
Fig. 22: Item map: pretest systemic thinking without i22
(level 1: light orange, level 2: medium orange, level 3: dark orange)

For the HLM analysis, the ArcGIS classes and students missing data in the age, sex or migration background variables are excluded, leading to a sample of 880 students (42 classes). Excluding also those not fitting for the binary family language background classification leaves 870 students (42 classes). The resulting "class" sizes range from eight to 28. For the analysis, WLE scores are used.

		coefficient	p
intercept		6.61	n.s.
variance components		variance component	p
u0		353.91	0.004
R		10177.03	
intraclass correlation	0.03	reliability estimate	0.42
deviance (parameters)		10512.16 (2)	
homogeneity test		n.s.	

Tab. 3: HLM: intercept only model

Looking at the intercept only model, based on the interpretation of SSI central (n.d.-b), only 3% of the variance is at class, the remaining 97% at student level (Tab. 3).

5.7.3 Conclusions

The study shows clearly the difficulty inherent in coding open ended tasks, especially due to the high variety of students' responses due to the larger sample size. The coding scheme presented here has undergone a development and refinement process, e.g. through the help of having an experienced teacher code a percentage of the sample with a preliminary system and then discuss the results to further improve the schema, discussions in expert rounds etc. These processes aim at making the coding as clear and logical as possible. Yet, there are several boundary cases remaining, potentially altering the results of the study depending on what classification is chosen in each case. Due to resource constraints, no double coding of the whole sample with the coding scheme used here was possible. For future work with similar resource constraints, closed tasks and as far as possible uniform task formats across levels seem advisable to avoid these problems.

Moreover, because even in the multiple choice items, it is sometimes difficult to decide what was checked and what was a previously made choice crossed out, clearer guidelines or computer based assessment should be contemplated. These guidelines could be similar, for instance, to high stakes tests, where only a specific way of checking is counted. These options have also potential draw-backs such as the need to access the school's computer room, increased reading time demand or not coding of student answers even if marked, but not in the specified way.

A more systematic test regarding the connection between linguistic skills and the performance on tests in systemic thinking in geography might also be valuable for the future. Furthermore, the test could be examined with further statistical measures, such as CFAs and differential item functioning.

5.8 Limitations due to the assessment instruments used

Due to the state of research in geography education, assessment instruments for systemic and spatial thinking had to be developed. These, naturally, have not been exposed to the same amount of scrutiny and sample sizes as normed instruments, especially compared to the time, samples and manpower used in large scale test development. Consequently, the results can only be first indications and need to be replicated by more refined assessment instruments in order to proof that the results hold up.

6 Describing the learning unit

Basis for the learning unit development is the choice of a content topic. One or several could be chosen. While a comparison of the impact of GIS by topic is of scientific interest, for this study, it would also increase the sample size required substantially. Moreover, such a comparison would require additional resources not only to develop two or more suitable tests and learning units, but also for instance to validate equivalency of tests and learning unit difficulties. Consequently, it was decided to focus on only one content topic.

The content topic was chosen based on several criteria. Firstly, it was important that the topic could be covered both with GIS and traditional paper maps, that is, that the necessary GIS data was available. Secondly, the topic needed to be suitable to the exploration of system relationships and allow for linking human- and physical-geographical subsystems. Thirdly, it should be a topic that can be related to more than one federal state curriculum for grade seven or the period including it, either through explicit inclusion or through open possibilities, for the high and mid stream (MKBB, 2002; MKBL, 2006; MKBR, 2006a, b; MKBW, 2004a, b; MKBY, 2001, 2004; MKHE, n.d.-a, b; MKHH, 2003, 2004; MKMV, 2002a, b; MKNS, 2008a, b; MKNW, 1993, 2004; MKRP, n.d.; MKSA, 1999, 2003; MKSH, 1997; MKSL, 2002, n.d.; MKSN, 2004a, b; MKTH, 1999a, b) and ideally be also relevant to the national educational standards (DGfG, 2007a). 'Tourism in Kenya' appeared to be one of the topics satisfying these criteria. Fourthly, German geography interest research as well as experts questioned informally suggested that from the available topics, 'tourism in Kenya' would be one that is potentially interesting. Especially girls have been shown to have an interest in people in other countries and developing countries (e.g. Hemmer & Hemmer, 2002a; Obermaier, 2002a). As girls on average have a significantly lower interest than boys in working with GIS or the computer in general (Klein, 2007), it was important to choose a topic/region that would interest them. Overall, Hemmer & Hemmer (2002a) have shown a medium interest of students in Africa. Consequently, the topic 'tourism in Kenya' was chosen for the treatment.

Kenya has received public attention for instance through the books and associated movies of Stefanie Zweig (e.g. Link, 2001; Zweig, 1995; 2000) and Corinne Hofmann (e.g. Hofmann, 2013), through the Nobel Peace Prize of Wangari Maathai (e.g. Nobel Foundation, 2004), through advertisement as well as through news coverage, for instance of the violence connected with the elections in 2007 (e.g. jba/AP/AFP, 2007; mik/AP/Reuters/dpa, 2007). Kenya has already been a destination for travel for more than 100 years (Akama, 1999). With rising tourist numbers over the last decades, tourism, which in Kenya is based both on natural and cultural assets, has had both positive and negative impacts on the country's environment and people in a complex network of interrelationships that also includes other factors such as population growth and agricultural developments. A brief overview over some of the foundational assets and resulting effects, which serve as base for the unit, can be found in the appendix. Besides this overview over the system, the materials were created after looking through existing materials and cannot be seen independently from test instrument creation and the accompanying survey of systemic thinking, spatial thinking and GIS studies (see ch. 5). Ideas can be taken e.g. both from school books (e.g. Baumann, et al., 2004; Obermann, 2005; Weidner, et al., 2006) and other materials (e.g. Albrecht, 1998; Buske, 2004; Clough & Holden, 2002; Dawson, n.d.; Dissen, 1990; Duke & Kolvoord, n.d.; Gaumnitz, 1993; Kaminske, 1996; Kerski, 2006; Learning Africa, n.d.; Malone, 2003; Malone & Wilson, n.d.; Meyer, 1995; Nölker, 1988; Sturm, 1981; Vettiger, et al., 2006). The tasks were meant to at least somewhat resemble the kind of tasks contained on worksheets and schoolbooks that teachers are offered. These include, for instance, materials provided on pages accompanying educational WebGIS such as those found on www.webgis-schule.de, special GIS books (e.g. Palmer, Palmer, Malone, & Voigt, 2008a, b; Palmer, Palmer, Malone, & Voigt, 2008c, d; Püschel, Hofmann, & Hermann, 2007), or school books (e.g. Obermann, 2005; Weidner, et al., 2006). For additional data sources used in the unit, see the appendix.

Drafts of the materials were discussed with experienced teachers.

6.1 Time frame

The time frame of the unit is an important consideration. Looking at curricular regulations, the curriculum of the federal state of Saxony suggests four lessons as a time frame for the unit entitled "Kenya" (translated), which is one of the compulsory elective units for *Mittelschule* (a school type including low and mid stream) in grade seven and focuses on tourism (MKSN, 2004a, p. 16). While existing GIS implementation studies often allow for a longer duration of the treatment (e.g. Aarons, 2003; Baker & White, 2003; Kerski, 2000), four lessons was judged to be a time frame that would provide a chance to notice an improvement in competencies and to have a greater limit on the number of confounding variables than longer treatments. A short treatment seems especially interesting for the study since it is realistic with regard to how GIS seems to be currently often used in German classroom practice, based on school books. For instance, Weidner et al. (2006) starts to introduce WebGIS on page 114. The introduction is, in total, four pages long. After that, no other task in the book seems to include the usage of GIS. This points to a rather short period in which GIS is used, based on a conversation with a teacher roughly about four lessons. Beyond these considerations, a rather short treatment period was considered conducive to finding a large enough number of participating teachers/classes (see ch. 3.4).

6.2 Basic methodological considerations

In contrast to the study of Ossimitz (1996), it was decided to provide the teachers with materials, and not to let them construct materials with the respective media on their own. Teachers are provided – online, in books or in teacher training courses – with pre-fabricated materials such as work sheets. A study by Baker, Palmer & Kerski (2009) shows that 68% of the sample, which includes 186 teachers that had participated in a GIS professional development course, uses the same materials in their teaching that had been used in the course. Another study (Kapulnik, Orion, & Ganiel, 2004), conducted with 202 teachers that participated in a three-year professional development course, also shows that if teachers implement new materials after the course, they are mostly those used in training. Nevertheless, providing the teachers with mate-

rials not only enhanced comparability on the student level and the possibility for statistical conclusions, but thus also could be fairly close to school practice. Providing the teachers with materials does have a drawback, however, as it introduces a new confounding variable. That the match between the provided material and the 'standard way' the classroom teacher teaches could potentially have a huge impact on the results came up as a point in a discussion with a teacher participating in study 5.

Studies have shown large instructor effects (see ch. 2). Since the sample size required to detect differences between two groups is fairly large, especially if the effect size is not big, it seemed not possible to have only one instructor for both groups. Thus, it was decided to allow for different instructors, but to at least somewhat minimize instructor effects by designing the materials for the unit in a way that the students can work by themselves.

An additional practical constraint is that in many schools, the number of computers in one lab is not sufficient for all students to be able to work alone. Consequently, in study 5 it was decided to let the students work in pairs. To ensure comparability, i.e. to not confound GIS vs. paper maps with pair work vs. no pair work, also the paper map students were asked to work in pairs. However, in both groups this also institutes a limitation with regard to the interpretation of the results (e.g. possible influences of 'pair chemistry'). In study 3, map and GIS students were separated in different rooms. In study 5, always the whole class was designed as either atlas or GIS in order to separate the groups (no influence of 'peeking' at the other medium) and because computer rooms often do not feature extra tables that would have allowed students to work with the atlas in the same room.

The materials were developed so that they were as far as possible parallel in both groups. This increases comparability, as in non-parallel materials, it would be unclear how much of the effect difference on achievement is due to the different tasks and how much is due to the different methods (GIS vs. atlas). However, it also introduces limitations, as many tasks where GIS could really show its strength can not, or only with great investment in time and effort, be answered with paper maps.

6.3 Study 3

For study 3 the materials were set to give the students the chance to explore the topic individually and provide an opportunity for authentic learning. For that, for instance the "[...] steps of geographic inquiry [...]" (ESRI, 2003) can provide some inspiration. The tasks were fairly open (in some aspects similar to e.g. Malone & Wilson, n.d.). A writing task example can be also found e.g. in Dawson (n.d.). The materials started with a brief description of a scenario about a school newspaper having a series "What our graduates are doing to-day", which is planned to feature several articles on two graduates wanting to open a travel firm specializing in nature-based tourism in Kenya. After students gained an overview about the available data (GIS or maps), the tasks led students through naming factors that are potentially relevant to judge the attractiveness of a plan in Kenya for such kind of tourism (e.g. closeness to animals for safaris) and exploring the distributions of these factors. Then students had to write an article for a school's newspaper including a brief situational description, three places especially interesting for nature-based tourism in Kenya, including an argument why these are interesting and a map underscoring the key arguments as well as a brief conclusion (product-orientation). Thus, the obligatory part of the materials focused mainly on parts of the assets for tourism in Kenya. The supplementary task was a second article asking for additional exploration of the student-selected places (e.g. with regard to the social situation), possible social, ecological and economical effects and conflicts resulting from increased tourist activity in this place as well as how the graduate's firm could respond to them in terms of sustainable travel. The article should again be supported by a suitable map.

The study was conducted in one day. The students started with the pretest, then worked on the learning materials, and then filled out the posttest. The author was present in the school when the study was conducted.

The GIS group worked with the GDV spatial commander. This was chosen as at that time, there was no WebGIS service for tourism in Kenya online yet, and the school did not have an ArcGIS license. The GDV spatial commander is a free DesktopGIS that seemed to be fairly popular in German GIS teacher train-

ing at the time. The atlas group worked with a school atlas and some additional paper maps (adapted from World Resources Institute, Department of Resource Surveys and Remote Sensing Ministry of Environment and Natural Resources Kenya, Central Bureau of Statistics Ministry of Planning and National Development Kenya, & International Livestock Research Institute, 2007, p. 14, p. 66, p. 83, p. 86) that focused specifically on Kenya.

In general, the students seemed engaged in the topic. However, in some cases, it seemed that the timeframe was tight although the compulsory materials only dealt with some factors for selecting places that are attractive for tourists. This can also be seen in the materials, were the supplementary article seems to not have been submitted by any of the students. Some seemed even to have troubles to complete the first article. For some students, the open tasks and the having to write a fitting article for the final product appeared to be challenging, and especially for the GIS group, exporting a map and printing it was often not easy. But also the map group had challenges, for instance, not all submitted maps featuring a legend.

The students for study 3 were a far more advanced sample then what was envisioned for the larger study 5 (twelfth vs. seventh graders). Consequently, it was decided to restructure the materials, making the format more guided and compress it in order to be able to include both the exploration of Kenya's assets and the effects of tourism. While this makes it easier for the students (and teachers), it does have costs with regard to fostering inquiry skills. It was also decided to leave out tasks that required students to create and print their own map. This helps in simplification and also reduces the IT support required, making it possible to conduct the unit in a wider range of schools.

6.4 Study 5

The basic structure of the materials for study 5 was adapted from Baumann et al. (2004, p. 134), a class seven textbook. It starts with some Kenya basics, such as location, area, population and natural conditions, then deals with tourism destinations. Afterwards, the chances and problems of tourism are discussed.

Besides by this basic structure and the general considerations outlined above, the development was also guided by constructing individual tasks based on a GIS methods competency model (see Fig. 23). Methods competencies and the partially overlapping concept of media competencies in general have been defined in a variety of ways (cp. e.g. DGfG, 2007a; Haubrich, 2001; Klein, 2007; Rinschede, 2003). No comprehensive empirically validated structure model of geographic methods competency (GMC) seemed to exist. There were first independent models (e.g. Bullinger & Hieber, 2004; Schubert & Uphues, 2008) or scoring guides (e.g. examples in Marcello, 2009) for different aspects of GMC, as well as considerations in how far the PISA reading literacy model can be used as a foundation for aspects of GMC (e.g. Hüttermann, 2004; Hüttermann, 2005; Lenz, 2004; Siegmund, Viehrig, & Volz, 2009). More recently, these have been extended e.g. by the theoretical development and empirical exploration of a model of satellite image reading literacy (Kollar, 2012), several theoretical and/or empirical works on an array of map competencies (e.g. Gryl & Kanwischer, 2012; Hemmer, Hemmer, Hüttermann, & Ullrich, 2010, 2012; Hemmer, Hemmer, Kruschel, et al., 2010; Hemmer, Hemmer, Kruschel, et al., 2012; Hüttermann, 2012; Hüttermann, Fichtner, & Herzig, 2010; Lenz, 2012; Lindau, 2010) and experimentation competencies (e.g. Otto, Mönter, & Hof, 2011).

Numerous GIS skill/skill progression/competency models have been proposed (cp. e.g. Board on Earth Sciences and Resources, 2006; Crechiolo, 1997; de Lange, 2007; Herzig, 2007; Hoenig & Niedenzu, 2004; Joachim, 2007; Johansson, 2006; O'Connor, 2005; Püschel, 2007; Püschel, et al., 2007; Schubert & Uphues, 2008; Treier, 2006). Many of them have been written for a specific target group, which can range from school students to university students preparing to be GIS professionals. Ultimately, it would be helpful to have one empirically and theoretically validated comprehensive competency structure model for the whole range from absolute beginner to GIS professional that is detailed enough to allow for differentiated classification and sense of achievement and yet generalized enough to be manageable. For this study, due to the still low level of implementation of GIS in German secondary schools, the materials focused primarily on basic GIS skills.

A first version of the model underlying the materials had been developed together with Dipl.-Geoökol. Daniel Volz and Prof. Dr. Alexander Siegmund (see Siegmund, et al., 2009; Volz, Viehrig, & Siegmund, 2010, also for further sources). For the study, it has been adapted and expanded. The basic structure of the expanded version is based on Einhaus (2007). The components for the dimension 'scale' are taken from geography's national educational standards (DGfG, 2007a). The dimension 'task complexity' is adapted from PISA's reading literacy (OECD, 2006a). The dimension 'media complexity' is included due to the many GIS didactical concepts that differentiate between Web- and DesktopGIS (e.g. Falk & Schleicher, 2005; Herzig, 2007; Püschel, 2007; Schäfer, 2007) and the often stated greater complexity of DesktopGIS making it more difficult to use (e.g. Baker, 2005). Even within one class of software (e.g. WebGIS), however, user interface design can make it easier or harder to retrieve information. In contrast, the dimension 'geodata access' deals not with the user interface but with what the students have to do in order to access the geodata to work with. The dimension 'background theory' is included due to the necessity of having a basic understanding of analysis possibilities in order to be able to efficiently retrieve information. Moreover, an extra division for background theory can be also found for instance in O'Connor (2005), or, for the method of programming, in Kohl (2008). Also important when talking about aspects of GMC are models about scientific thinking (e.g. Grube, Möller, & Mayer, 2007; Mayer, 2007) and models of dealing with data and graphical representations (e.g. Kuntze, Lindmeier, & Reiss, 2008; Lindmeier, Kuntze, & Reiss, 2007).

Additionally, several other tasks are included. Two of them were very basic tasks requiring students to retrieve information from simple tables. As in a GIS, the attribute tables and the graphic representation are interlinked, inability to retrieve information from tables would likely lead to great difficulties mastering GIS tasks. Furthermore, some contextual tasks such as requiring students to think about their pre-knowledge of Kenya and their pre-experience with the medium, a task analyzing photos from Kenya and some text analysis tasks were included, in order to create an as far as possible 'round' unit despite the limitations that are inherent in the study design. To further improve the materials, they were discussed with experts, such as experienced teachers.

In the course of conducting the main study it was decided to introduce a second format, based on the feedback provided by the first parts of the sample. In contrast to the first format where the GIS tools are explained on a separate sheet ('old' materials), the explanations how to do something in GIS are directly included on the worksheets close to the tasks where they are required, and are much more 'step-by-step' ('new' materials). Moreover, the layout is slightly changed. The tasks themselves, however, remain parallel as far as possible.

dimensions	components				
content area	population	vulcanoes	Kenya tourism	...	
scale	local	regional	national	international	global
task complexity	I locate information which typically only has to satisfy one criterium				
	II locate information which has to satisfy several criteria; compare or combine two pieces of information				
	III compare or combine several pieces of information				
	...				
media complexity	I low				
	II intermediate				
	III high				
	...				
background theory	I basic, general idea				
	II basics of symbology and visualization				
	III basics of analysis and logic operators				
	...				
geodata access	I open given ready-to-use GIS projects				
	II select geodata from a given structure				
	III independently aquire geodata from different sources				
	...				

Fig. 23: Rubric for task construction
(adapted and expanded from Siegmund, et al., 2009)
components covered by the materials are marked in grey

The atlas group materials were created for the Diercke World Atlas (Westermann, 2008) and for some additional paper maps for Kenya (based on the GIS data and adapted from World Resources Institute, et al., 2007, p. 83). It was decided to focus only on one atlas to ensure comparability. Schools in Baden-Wurttemberg who did not have this atlas were provided with the required number for the duration of the study. For the GIS group, the materials were created in different versions. The first version was created for the Diercke WebGIS, which is accessible from any computer with internet access, needs no extra installations and has been designed specifically for use in school. The second version was created for the GDV Spatial Commander. In the study, however, none of the teachers chose to work with it. The third version was created for ArcGIS 9.3, a professional DesktopGIS software used in many research fields and companies. While on the one hand, due to its primary target group (professionals), the software is very complex, it is in many ways also easier to retrieve information from it than from educational WebGIS (for instance, the complete attribute table can be shown and sorted, allowing also the selection of an entry in the table that is then highlighted in the map; users can themselves easily adjust the ordering and transparency of the layers, etc.). Nevertheless, only two of the participating classes used ArcGIS.

The GIS data for the materials was collected from different sources (see appendix) and transformed where necessary to be fittingly matched and overlaid. Attributes were selected, translated and simplified in order to be suitable to the planned unit. The GIS data development was conducted together with Dipl.-Geoökol. Daniel Volz. Some technical issues, such as troubles with an originally planned inclusion of photos via hyperlinks had to be worked around (e.g. through using a photo transparency).

The student feedback varies greatly. Those students that answered the feedback question gave a diverse assessment of the unit, reaching from hard to easy, from boring, no fun or stupid to interesting, fun and good, as well as a broad spectrum in between.

Of the map students, only 1.2% specified that they hadn't worked with an atlas before (1.2% missing, n=326). In contrast, only 7.3% of the GIS students an-

swered that they had already worked with GIS (1.2% missing, n=563). When counting only those that actually specified an educational GIS, such as *Diercke* or *Webgis-Schule*, and not those that e.g. had a blank answer, satnav or Google Earth, the number decreases to 23 or 4.1%.

The differences in familiarity between the two media also show up in the feedback, although both groups contained students that thought the work with the respective medium was easy and those that though it was hard. Some students, especially in the GIS group, gave feedback such as ID 1055 "It was hard at the beginning but after some time it's been easy for me" (translated) or ID 0944 "At the beginning it was hard to deal with the GIS but later, when we have understood it, it was quite ok." (translated). Additionally, it is interesting to note that these difficulties with GIS do not seem to be a product of the age of students. Feedback from one class that specified to be in grade eleven, also included comments such as "Didn't like it at all, it took a long time to find something and the program is complicated" (ID 0850, translated) or "Not good. Because I have worked with the program for the first time, I had to familiarize myself [with it] a long time. Moreover I found the program very confusing and just hard to understand" (ID 0857, translated). Overall, this hints at that at least some of the students at the beginning needed to get used to the new medium. The imbalance in pre-experience, and thus likely methods competencies, could have a potentially large influence on the results (see ch. 7 and 8).

It also needs to be taken into account that technical systems are prone to bugs, and at least some of the WebGIS groups reported having troubles with the platform, such as not getting the attribute query to work properly in their school.

6.5 Limitations due to the learning units used

The treatment materials have a profound effect on the outcome of the study. Together with the assessment instrument used, the materials are thus the study's greatest potential limitation. In general, as the materials only cover one topic and one basic structure, the results are not generalizable to all GIS vs. atlas materials.

Based on the current situation of GIS education in Germany it was decided to focus on basic GIS lessons working with given data. Care was taken to design

the materials both similar to existing materials and based on theoretical considerations, a well as to revise them based on a first trial. However, time and curriculum constraints made it impossible to have materials for the study that had been refined through a large number of pilot runs. Furthermore, using the same materials for all in one particular group and letting the students work independently in pairs contributes to reducing instructor effects, increases the possibilities for statistical conclusions and models the not uncommon situation of using pre-fabricated worksheets. At the same time, however, it also introduces potential confounding variables and limits the conclusions that can be drawn. Additionally, it needs to be taken into account that the materials were developed based only on a survey of literature, a look at some existing teaching materials and short informal conversions with few people who had lived in Kenya, not own experience. Unfortunately, the development of the teaching materials coincided with the unrest following the elections in Kenya making traveling to the country not advisable.

7 Analyzing the findings

The five studies conducted within this dissertation produced a rich data set. However, it has to be kept in mind that the studies are exploratory in nature. The analysis will focus on direct effects as described in the hypotheses. As such, the focus will be on the two pre-/posttest studies, especially study 5. While the analysis is conducted both with raw and WLE values in many cases, graphics will focus on the results of the WLE-based analyses. All HLM results are reported based on robust standard errors.

7.1 Study 3

Study 3 is a pre-/posttest design study. Due to the variables included and the very small sample size not all hypotheses can be tested. An evaluation of the cell sizes showed that several would be problematic.

7.1.1 Hypothesis I: Improvement through GIS

I   On average, there will be an improvement in systemic thinking achievement for participants working with GIS in the unit.

The GIS group improves not significantly (Fig. 24, Tab. 4). However, both effect sizes are above 0.2. Consequently, hypothesis I can only be partly accepted.
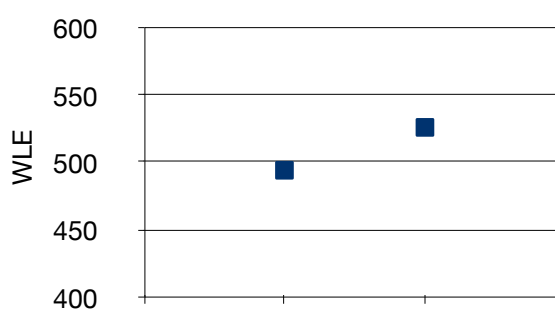
Fig. 24: Pre- and posttest mean values GIS group

	raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
	M	**SD**				**M**	**SD**				
pre	0.58	0.19	0.07	0.35	0.09	494.02	112.82	0.05	0.39	0.07	31
post	0.63	0.23				525.84	116.81				

Tab. 4: Pre- and posttest mean values (GIS group)

7.1.2 Hypothesis II: Improvement map vs. GIS

II On average, there will be a greater improvement in systemic thinking achievement for participants working with GIS in the unit than for participants working with paper maps.

The map group improves, too. The change has a greater effect size than the GIS group and both raw and WLE values are significant (Tab. 5). There are no significant differences between the two groups ($d<0.2$) (Fig. 25, Tab. 6, Levene's test WLE $p=0.06$).

	raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
	M	**SD**				**M**	**SD**				
pre	0.62	0.15	0.008	0.50	0.008	505.79	87.25	0.002	0.57	0.008	32
post	0.67	0.14				541.09	70.41				

Tab. 5: Pre- and posttest mean values (map group)

	raw		p	|d|	p (MW)	WLE		p	|d|	p (MW)	n
	M	**SD**				**M**	**SD**				
map	0.05	0.10	n.s.	< 0.2	n.s.	35.30	60.36	n.s.	< 0.2	n.s.	32
GIS	0.05	0.14				31.83	86.79				31

Tab. 6: Difference score mean values
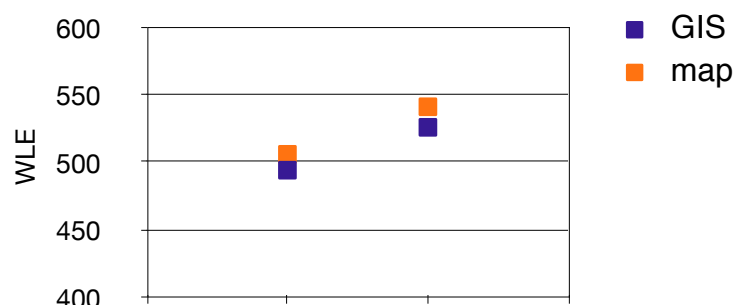
Fig. 25: Pre- and posttest mean values by group

Consequently, the hypothesis needs to be rejected (Tab. 7).

		significance	effect size
difference in prepost-change between GIS and map group	**(t)**	✘	✘
	(MW)	✘	

Tab. 7: Overview of tests for hypothesis II
("On average, there will be a greater improvement in systemic thinking achievement for participants working with GIS in the unit than for participants working with paper maps.") (✘: contrary to hypothesis, ✔: in line with hypothesis)

7.1.3 Hypothesis III-b: Improvement and sex

III On average ...

   b) ... there is no effect of sex on improvement.

Male map students' WLE score increases significantly. Female students improve significantly in the GIS group for both WLE and raw score and in the map group for the raw score. Nearly all effect sizes are above 0.2 (Tab. 8).

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	male	pre	0.59	0.15	n.s.	0.46	n.s.	479.70	84.38	0.007	0.91	0.01	13
		post	0.63	0.15				534.62	83.87				
	female	pre	0.63	0.16	0.02	0.58	0.03	522.43	89.16	n.s.	0.39	n.s.	18
		post	0.69	0.13				545.51	63.42				
GIS	male	pre	0.59	0.22	n.s.	0.22	n.s.	497.65	124.11	n.s.	< 0.2	n.s.	11
		post	0.56	0.24				491.87	117.59				
	female	pre	0.58	0.18	0.01	0.66	0.009	492.02	109.44	0.02	0.55	0.05	20
		post	0.66	0.22				544.53	115.03				

Tab. 8: Pre- and posttest mean values by group and sex

There are significant differences only for the GIS group, in favor of female students (*d*>0.2 in nearly all cases; Tab. 9, Fig. 26, Levene's test GIS WLE *p*=0.046). Splitting the sample by sex, only for the male students' WLE score the difference is significant (*d*>0.2 in all cases; Levene's test female WLE *p*=0.08). Thereby, for male students, working with maps is favorable and for female students working with GIS (Tab. 10, Fig. 27).
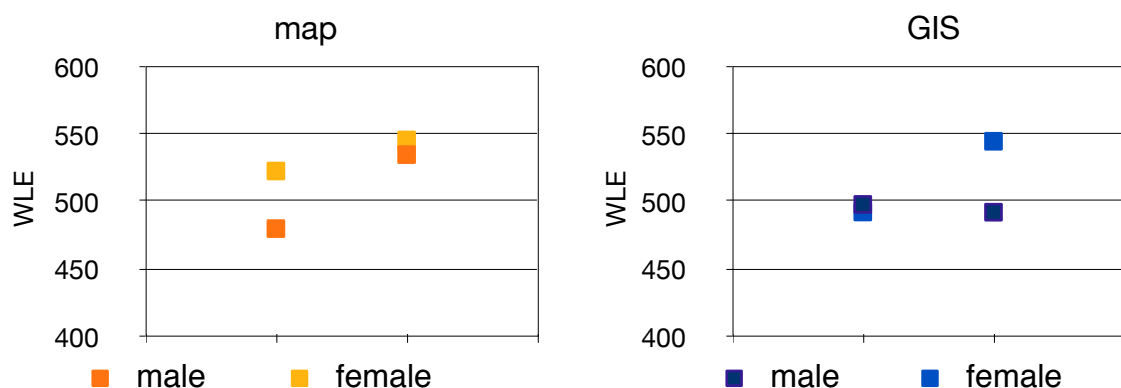
Fig. 26: Pre- and posttest mean values by group and sex

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	male	0.04	0.09	n.s.	< 0.2	n.s.	54.92	60.31	n.s.	0.53	n.s.	13
	female	0.06	0.10				23.08	59.58				18
GIS	male	-0.03	0.12	0.03	0.89	0.04	-5.78	55.84	0.04	0.75	n.s.	11
	female	0.09	0.13				52.51	94.77				20

Tab. 9: Difference score mean values by group and sex
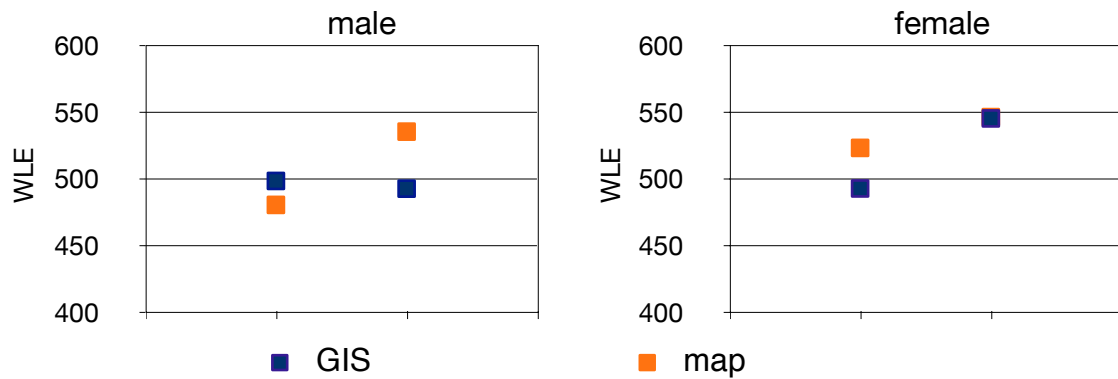
Fig. 27: Pre- and posttest mean values by sex and group

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
male	map	0.04	0.09	n.s.	0.65	n.s.	54.92	60.31	0.02	1.04	0.02	13
	GIS	-0.03	0.12				-5.78	55.84				11
female	map	0.06	0.10	n.s.	0.22	n.s.	23.08	59.58	n.s.	0.37	n.s.	18
	GIS	0.09	0.13				52.51	94.77				20

Tab. 10: Difference score mean values by sex and group

In two way ANOVAs (Levene's test raw n.s., WLE $p=0.06$), only the WLE-interaction and the raw score effect for sex are significant (Tab. 11, Fig. 28).

Fig. 28: Two way ANOVA graphical results: interaction group*sex

	raw				WLE			
	F	**p**	**adjusted R^2**	**partial η^2**	**F**	**p**	**adjusted R^2**	**partial η^2**
group	0.54	n.s.		0.01	0.68	n.s.		0.01
sex	4.57	0.04	0.06	0.07	0.49	n.s.	0.05	0.01
sex*group	2.61	n.s.		0.04	5.64	0.02		0.09

Tab. 11: Two way ANOVA: group, sex and interaction

The results are mixed (Tab. 12), hinting at male students faring better with maps and female students with GIS. Consequently, the hypothesis should be rejected.

		significance	effect size
prepost-change within the male and female sub-groups for the map group	**(t)**	✘	✔
	(W)	✘	
prepost-change within the male and female sub-groups for the GIS group	**(t)**	✘	✘ (WLE)/ ✔ (raw)
	(W)	✘ (raw)/ ✔ (WLE)	
difference in prepost-change between the sex sub-groups within the map group	**(t)**	✔	✘ (WLE) / ✔ (raw)
	(MW)	✔	
difference in prepost-change between the sex sub-groups within the GIS group	**(t)**	✘	✘
	(MW)	✘ (raw)/ ✔ (WLE)	
difference in prepost-change between the map and GIS group within the sub-groups	**(t)**	✘ (WLE)/ ✔ (raw)	✘
	(MW)	✘ (WLE)/ ✔ (raw)	
two way ANOVA with group, sex and interaction		✘	✘

Tab. 12: Overview of tests for hypothesis III-b
 ("On average, there is no effect of sex on improvement.")
 (✘: contrary to hypothesis, ✔: in line with hypothesis)

7.1.4 Hypothesis III-e: Improvement and pre-achievement

III On average, ...

   e)... pre-achievement has an effect on improvement.

The sample was split based on a pre-WLE score of above or below 500, with 36 students (57.1%) in the low and 27 (42.9%) in the high group. The analysis is thus conducted based on the WLE scores. For the low group, map and GIS students improve ($d > 0.2$), although results are mixed. With regard to high systemic thinking students no group improves significantly ($dz < 0.2$, Tab. 13).

WLE			M	SD	p	|dz|	p (W)	n
map	< 500	pre	439.96	45.37	< 0.001	1.14	0.001	18
		post	501.85	56.69				
	> 500	pre	590.44	41.74	n.s.	< 0.2	n.s.	14
		post	591.54	52.41				
GIS	< 500	pre	418.35	68.07	0.048	0.50	0.05	18
		post	463.95	97.38				
	> 500	pre	598.79	69.09	n.s.	< 0.2	n.s.	13
		post	611.54	83.53				

Tab. 13: Pre- and posttest mean values by group and pre-score (bin.)

Within one group (Fig. 29, Tab. 14), only the map students' differences are significant. Both groups have effect sizes above 0.2., in favor of low pre-score students. There is no significant difference within one pre-achievement group between map and GIS (Tab. 15, Fig. 30). For the below 500 group, the effect size just above 0.2, in favor of map students (Levene's test, p=0.04).

diff WLE		M	SD	p	|d|	p (MW)	n
map	< 500	61.89	54.24	0.003	1.15	0.008	18
	> 500	1.10	51.00				14
GIS	< 500	45.60	91.00	n.s.	0.38	n.s.	18
	> 500	12.75	80.12				13

Tab. 14: Difference score mean value by group and pre-score (bin.)
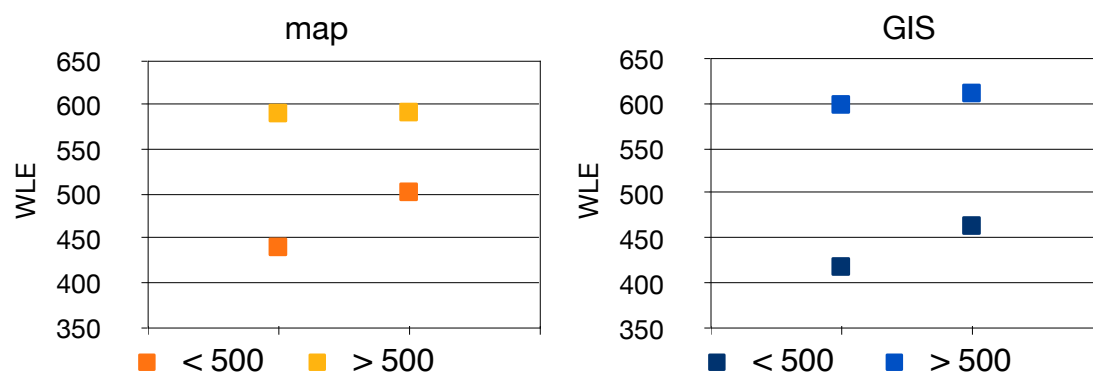
Fig. 29: Pre- and posttest mean values by group and pre-score (bin.)

diff WLE		M	SD	p	|d|	p (MW)	n
< 500	map	61.89	54.24	n.s.	0.22	n.s.	18
	GIS	45.60	91.00				18
> 500	map	1.10	51.00	n.s.	< 0.2	n.s.	14
	GIS	12.75	80.12				13

Tab. 15: Difference score mean value by pre-score (bin.) and group
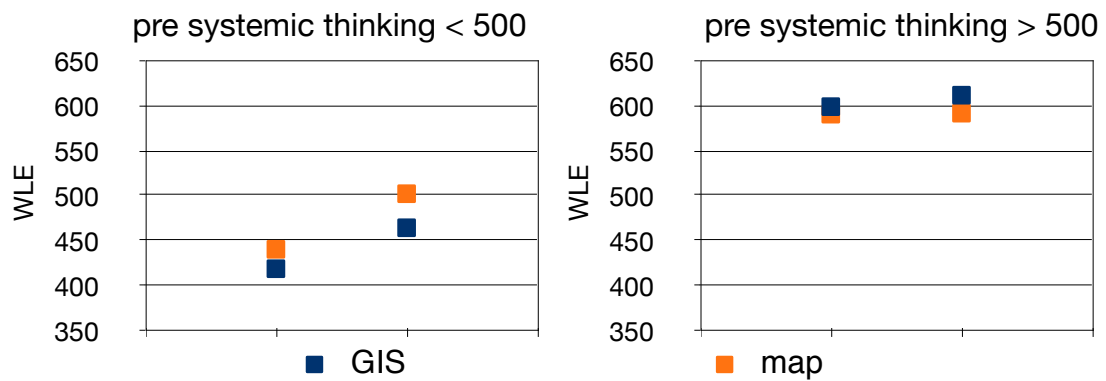
Fig. 30: Pre- and posttest mean values by pre-score (bin.) and group

Two way ANOVAs show only a significant effect for pre-score (bin.) (Levene's test p=0.06, Tab. 16, Fig. 31).
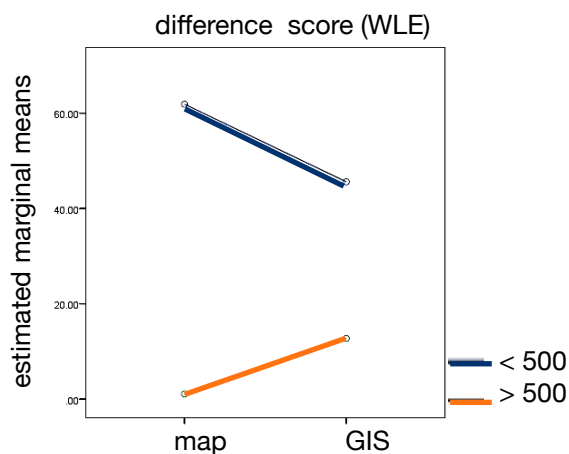
Fig. 31: Two way ANOVA graphical results: interaction group*pre-score

WLE	F	p	adjusted R^2	partial η^2
group	0.02	n.s.		0.00
pre-score bin.	6.61	0.01	0.07	0.10
pre-score bin.*group	0.59	n.s.		0.01

Tab. 16: Two way ANOVA: group, pre-score (bin.) and interaction

The results are mixed but mostly in favor of the hypothesis (Tab. 17).

		significance	effect size
prepost-change within the two sub-groups for the map group	**(t)**	✔	✔
	(W)	✔	
prepost-change within the two sub-groups for the GIS group	**(t)**	✔	✔
	(W)	✔	
difference in prepost-change between the two sub-groups within the map group	**(t)**	✔	✔
	(MW)	✔	
difference in prepost-change between the two sub-groups within the GIS group	**(t)**	✘	✔
	(MW)	✘	
difference in prepost-change between the map and GIS group within the sub-groups	**(t)**	✘	✔
	(MW)	✘	
two way ANOVA: group, pre-score (bin.) and interaction		✔	✔

Tab. 17: Overview of tests for hypothesis III-e
("On average, pre-achievement has an effect on improvement.")
(✘: contrary to hypothesis, ✔: in line with hypothesis)

7.1.5 Hypothesis III-f: Improvement and pre-spatial thinking

III On average, ...

   f) ... there is no effect of pre-spatial thinking score on improvement.

Due to the unfavorable statistical properties of the spatial thinking items, the possibilities for exploration seem questionable. To get an exploratory insight, the pre-spatial thinking score was coded binary (0=solved 1 or 2 items, 1=solved 3 or 4 items). There are 28 (44.4%) in the low and 35 (55.6%) in the high group. For the low spatial thinking group, only GIS students improve significantly (d>0.2), albeit the raw score map students have an effect size of 0.2. In the high group, only the map students improve significantly (d>0.2, Tab. 18).

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	**SD**				**M**	**SD**				
map	low	pre	0.59	0.15	n.s.	0.20	n.s.	501.14	78.38	n.s.	< 0.2	n.s.	10
		post	0.61	0.12				512.06	67.01				
	high	pre	0.63	0.16	0.006	0.65	0.007	507.00	92.74	0.001	0.87	0.003	22
		post	0.69	0.14				554.28	69.36				
GIS	low	pre	0.54	0.16	0.020	0.61	0.049	469.20	109.93	0.03	0.56	0.03	18
		post	0.62	0.22				517.18	107.07				
	high	pre	0.65	0.22	n.s.	< 0.2	n.s.	528.39	111.78	n.s.	< 0.2	n.s.	13
		post	0.64	0.25				537.85	132.67				

Tab. 18: Pre- and posttest mean values by group and pre-spatial thinking score (bin.)

Only the GIS raw score students' difference is significant ($d>0.2$ in all cases, Fig. 32, Tab. 19). Only the high group raw score difference is significant (Fig. 33, Tab. 20, WLE Levene's test $p=0.07$). The effect sizes are above 0.2 in all cases, with low group students faring better with GIS and high group ones with maps.

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	low	0.02	0.11	n.s.	0.40	n.s.	8.92	67.25	0.096	0.63	n.s.	10
	high	0.06	0.09				47.28	54.38				22
GIS	low	0.09	0.14	0.045	0.78	0.03	47.98	86.11	n.s.	0.45	n.s.	18
	high	-0.01	0.10				9.46	85.98				13

Tab. 19: Difference score mean values by group and pre-spatial thinking score (bin.)
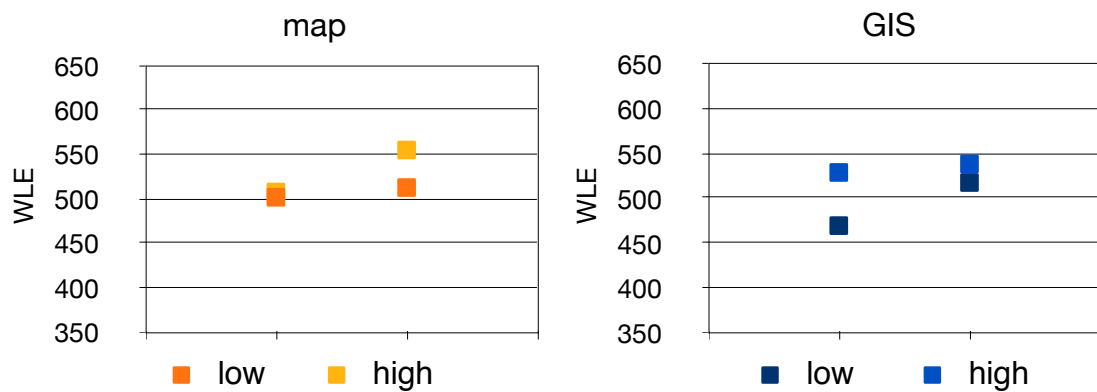
Fig. 32: Pre- and posttest mean values by group and pre-spatial thinking score (bin.)

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
low	map	0.02	0.11	n.s.	0.52	n.s.	8.92	67.25	n.s.	0.51	n.s.	10
	GIS	0.09	0.14				47.98	86.11				18
high	map	0.06	0.09	0.04	0.73	0.06	47.28	54.38	n.s.	0.53	n.s.	22
	GIS	-0.01	0.10				9.46	85.98				13

Tab. 20: Difference score mean values by pre-spatial thinking score (bin.) and group
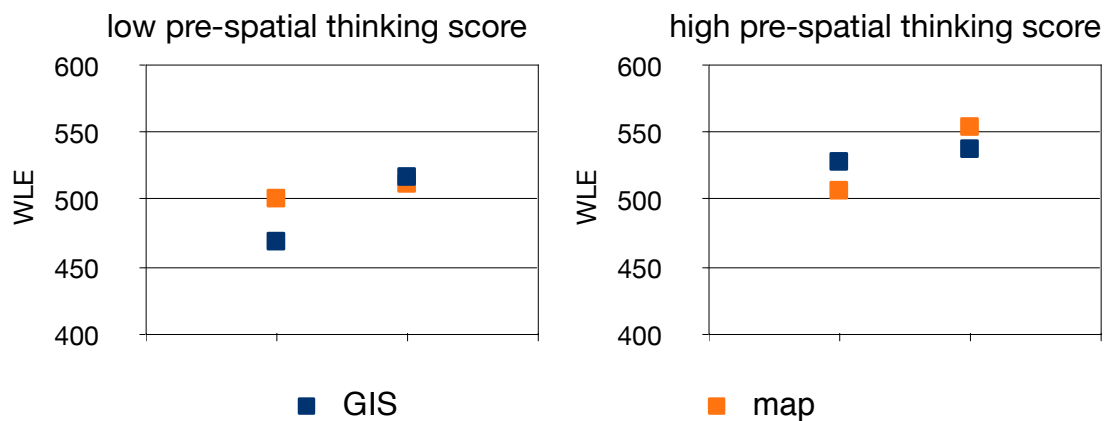
Fig. 33: Pre- and posttest mean values pre-spatial thinking score (bin.) and group

In the two way ANOVAs, only the raw score interaction is significant (Levene's test n.s., Fig. 34, Tab. 21). However, in both analyses, the interaction has comparatively large effect sizes.

difference score (WLE)
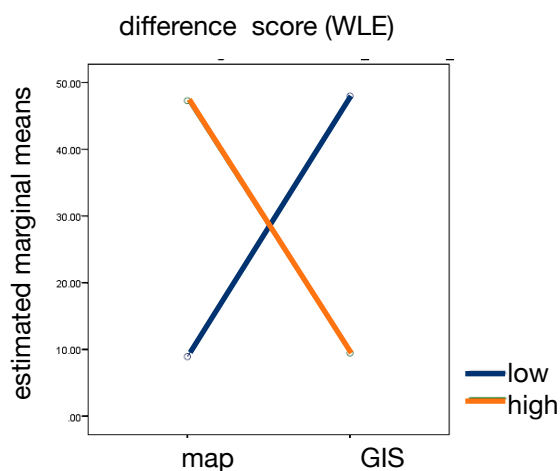
Fig. 34: Two way ANOVA graphical results: interaction group*pre-spatial thinking score (bin.)

	raw				WLE			
	F	**p**	**adjusted R^2**	**partial η^2**	**F**	**p**	**adjusted R^2**	**partial η^2**
group	0.01	n.s.		0.00	0.00	n.s.		0.00
pre-spat. bin.	0.93	n.s.	0.05	0.02	0.00	n.s.	0.02	0.00
pre-spat. bin.*group	5.27	0.03		0.08	3.95	0.05		0.06

Tab. 21: Two way ANOVAs: group, pre-spatial thinking score (bin.) and interaction

		significance	**effect size**
prepost-change within the two sub-groups for the map group	(t)	✗	✗
	(W)	✗	
prepost-change within the two sub-groups for the GIS group	(t)	✗	✗
	(W)	✗	
difference in prepost-change between the two sub-groups within the map group	(t)	✔	✗
	(MW)	✔	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✗ (raw)/ ✔ (WLE)	✗
	(MW)	✗ (raw)/ ✔ (WLE)	
difference in prepost-change between the map and GIS group within the sub-groups	(t)	✗ (raw)/ ✔ (WLE)	✗
	(MW)	✔	
two way ANOVA: group, pre-spatial thinking score (bin.) and interaction		✗ (raw)/ ✔ (WLE)	✗

Tab. 22: Overview of tests for hypothesis III-f
("On average, there is no effect of pre-spatial thinking score on improvement.")
(✗: contrary to hypothesis, ✔: in line with hypothesis)

The results for this hypothesis are mixed, but largely against the hypothesis (Tab. 22). Consequently, the hypothesis should be rejected.

7.1.6 Conclusions from Study 3

First results show that overall, GIS is not advantageous compared to maps, although also GIS students improve pre- to posttest. There are some differences in achievement change based on student characteristics, albeit often contrary to expectations. The inclusion of spatial thinking as a variable is reasonable. However, the sample size is very small, thus the results are to be interpreted very cautiously.

7.2 Study 5

Study 5 is a pre-/posttest-design study. Thus, all hypotheses can be tested.

7.2.1 Hypothesis I: Improvement

I   On average, there will be an improvement in systemic thinking achievement for participants working with GIS in the unit.

GIS group mean values stay nearly the same (Fig. 35, Tab. 23, 24). Over all GIS classes 36.7% of the students decrease, 27.2% stay the same and 36.2% increase in their achievement.
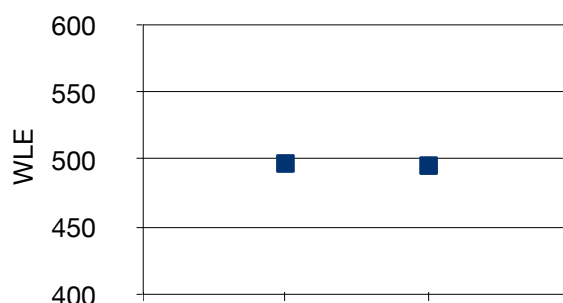
Fig. 35: Pre- and posttest mean values for the GIS group

	raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
	M	SD				M	SD				
pre	0.47	0.22	n.s.	< 0.2	n.s.	496.99	99.51	n.s.	< 0.2	n.s.	603
post	0.47	0.25				495.43	114.95				

Tab. 23: Pre- and posttest mean values (GIS group)

	raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
	M	SD				M	SD				
pre	0.47	0.22	n.s.	< 0.2	n.s.	495.64	98.60	n.s.	< 0.2	n.s.	574
post	0.47	0.25				494.55	116.29				

Tab. 24: Pre- and posttest mean values (GIS group without ArcGIS classes)

Consequently, the hypothesis needs to be rejected.

7.2.2 Hypothesis II: Improvement GIS vs. map

II On average, there will be a greater improvement in systemic thinking achievement for participants working with GIS in the unit than for participants working with paper maps.

For the map group, achievement improves significantly (Tab. 25). Over all map classes 28.0% of the students decrease, 27.7% stay the same and 44.4% increase in their achievement. Tests show that the difference between the GIS and map group is significant, but has effect sizes below 0.2 (Fig. 36, Tab. 26, 27).
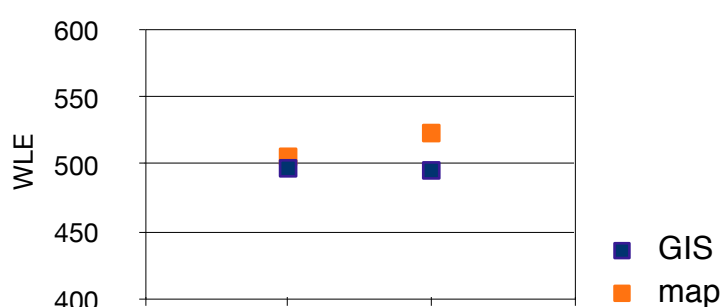
Fig. 36: Pre- and posttest mean values by group

	raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
	M	SD				M	SD				
pre	0.49	0.23	0.001	< 0.2	0.001	505.52	100.81	0.002	< 0.2	0.005	329
post	0.53	0.25				523.15	117.36				

Tab. 25: Pre- and posttest mean values (map group)

	raw		p	|d|	p (MW)	WLE		p	|d|	p (MW)	n
	M	SD				M	SD				
map	0.04	0.23	0.007	< 0.2	0.004	17.63	102.70	0.006	< 0.2	0.006	329
GIS	0.00	0.22				-1.56	101.66				603

Tab. 26: Difference score mean values by group

	raw		p	|d|	p (MW)	WLE		p	|d|	p (MW)	n
	M	SD				M	SD				
map	0.04	0.23	0.009	< 0.2	0.005	17.63	102.70	0.008	< 0.2	0.007	329
GIS	0.00	0.22				-1.09	101.87				574

Tab. 27: Difference score mean values by group (without ArcGIS classes)

In a HLM, GIS as a level 2 predictor has a significant negative effect (Tab. 28). 22.8% of the class variation can be explained by the treatment medium. However, the intercept chi-square is significant, indicating that there is still class difference

score variation that "[...] remains to be explained" (UCLA Statistical Consulting Group, n.d.).

	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	19.81	0.002	273.10	10174.76	0.016	0.23
GIS	-20.78	0.02				
deviance (parameters)	10502.54 (2)		homogeneity test	n.s.	reliability	0.35

Tab. 28: HLM: GIS as level 2 predictor model

GIS students' pre-posttest change is significantly worse than map students'. Consequently (Tab. 29), the hypothesis needs to be rejected.

		significance	effect size
difference in prepost-change between GIS and map group	(t)	✗	✗
	(MW)	✗	
HLM		✗	

Tab. 29: Overview of tests for hypothesis II
("On average, there will be a greater improvement in systemic thinking achievement for participants working with GIS in the unit than for participants working with paper maps.") (✗: contrary to hypothesis, ✔: in line with hypothesis)

7.2.3 Hypothesis III-a: Improvement and age

III Improvement of achievement: On average, ...

   a) ... there is no effect of age on improvement.

Looking at the distributions in the map and GIS groups individually, the ages eleven, 15, and 16 each have less than five students. The age group 16 is only represented in the GIS group. Consequently, these students are excluded for all analyses except the HLM.
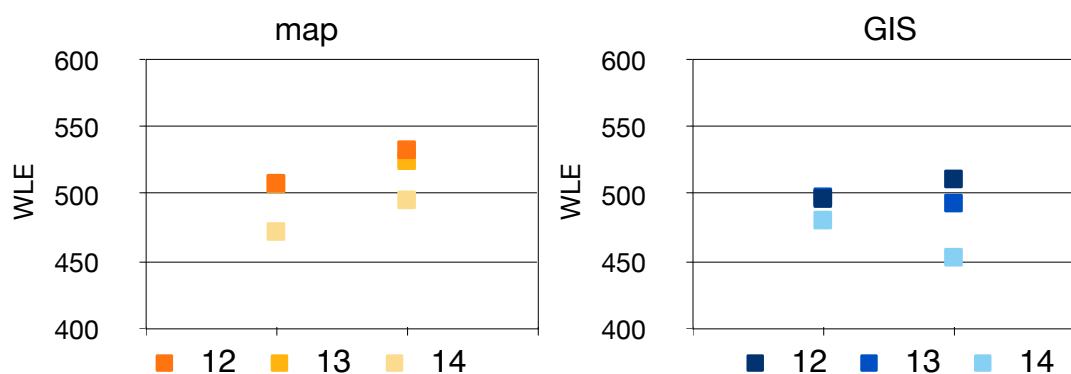
Fig. 37: Pre- and posttest mean values by group and age (selection)

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	12	pre	0.50	0.23	0.005	0.29	0.007	507.99	96.09	0.01	0.26	0.03	100
		post	0.56	0.22				532.86	97.64				
	13	pre	0.49	0.23	0.01	< 0.2	0.01	507.56	104.22	0.02	< 0.2	0.02	208
		post	0.53	0.26				524.72	121.43				
	14	pre	0.41	0.17	n.s.	0.26	n.s.	472.16	70.37	n.s.	< 0.2	n.s.	16
		post	0.47	0.29				495.65	143.23				
GIS	12	pre	0.47	0.23	0.08	< 0.2	0.09	496.61	101.24	0.08	< 0.2	0.09	145
		post	0.50	0.25				511.16	113.16				
	13	pre	0.47	0.23	n.s.	< 0.2	n.s.	497.92	99.47	n.s.	< 0.2	n.s.	405
		post	0.47	0.25				493.27	116.31				
	14	pre	0.44	0.22	0.04	0.33	0.04	480.65	101.06	0.06	0.31	0.06	41
		post	0.38	0.22				453.05	102.66				

Tab. 30: Pre- and posttest mean values by group and age (selection)

The results show a differentiated picture, with significance and effect sizes above 0.2 not being achieved in all sub-groups (Fig. 37, Tab. 30). Overall, it is noticeable that in both the map and the GIS group, the 12-year-olds show the highest posttest score. It is also noteworthy that the 14-year-olds have the lowest pretest score in both GIS and map group. In the map group, there is no significant change (but the increase in the raw score has a $dz>0.2$). In the GIS group, there is a decrease in achievement ($dz>0.2$) for this subgroup, which is significant for the raw but not for the WLE score.

There is no significant difference by age for the map, but for the GIS group, (Fig. 38, Tab. 31, ANOVA Levene's test n.s.). Even for the GIS group, however, effect size is below RMPE.
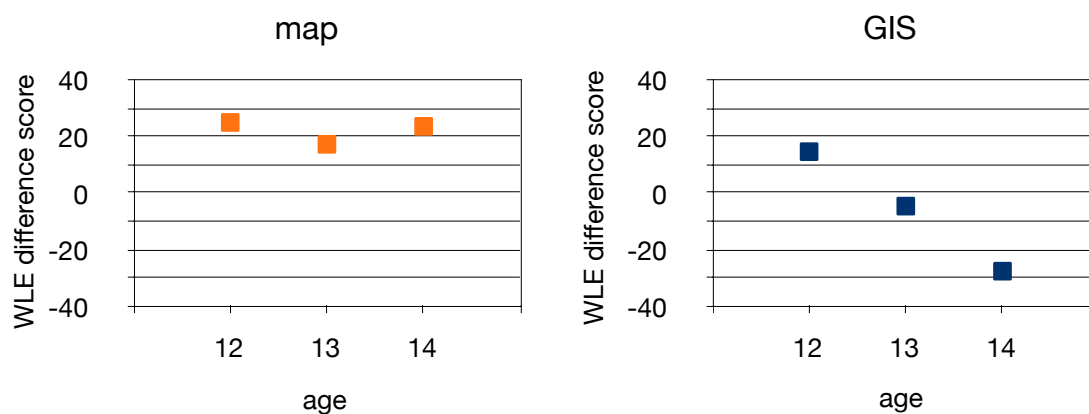
Fig. 38: WLE difference score mean values by group and age (selection)

		diff raw		ANOVA raw				KW	diff WLE		ANOVA WLE				KW	
		M	SD	F	p	ad-just. R^2	par-tial η^2	p	M	SD	F	p	ad-just. R^2	par-tial η^2	p	n
map	12	0.06	0.21						24.87	95.68						100
	13	0.04	0.22	0.38	n.s.	-0.004	0.002	n.s.	17.16	101.97	0.21	n.s.	-0.01	0.001	n.s.	208
	14	0.06	0.24						23.49	123.14						16
GIS	12	0.03	0.22						14.56	99.15						145
	13	-0.00	0.23	3.29	0.04	0.01	0.01	0.04	-4.65	104.00	3.34	0.04	0.01	0.01	0.04	405
	14	-0.07	0.20						-27.60	90.07						41

Tab. 31: Difference score mean values by group and age (selection)

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
12	map	0.06	0.21	n.s.	< 0.2	n.s.	24.87	95.68	n.s.	< 0.2	n.s.	100
	GIS	0.03	0.22				14.56	99.15				145
13	map	0.04	0.22	0.03	< 0.2	0.03	17.16	101.97	0.01	0.21	0.02	208
	GIS	-0.00	0.23				-4.65	104.00				405
14	map	0.06	0.24	0.045	0.58	0.05	23.49	123.14	0.09	0.47	n.s.	16
	GIS	-0.07	0.20				-27.60	90.07				41

Tab. 32: Difference score mean values by age (selection) and group

Comparing map and GIS students within one age shows differentiated results (Fig. 39, Tab. 32). There is no significant difference in the increase ($d<0.2$) between 12-year-old map and GIS students. However, Fig. 39 highlights that GIS students' pre-posttest-change seems to get progressively worse in the older age subgroups.
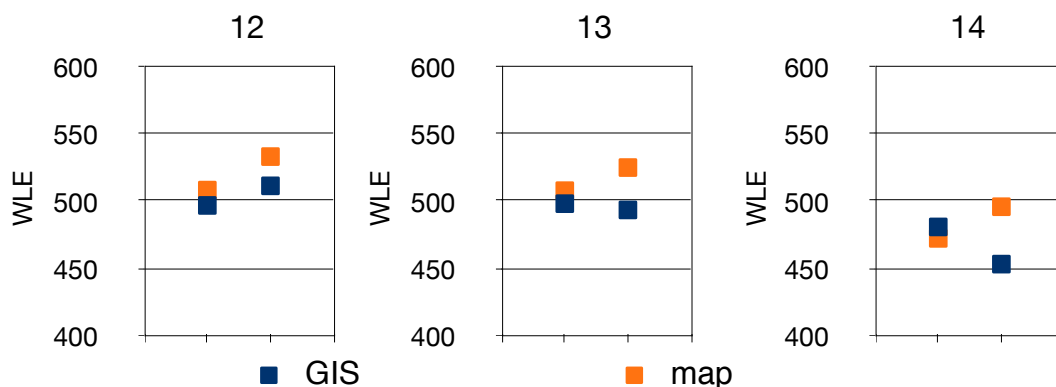
Fig. 39: Pre- and posttest mean values by age and group (selection)

		age*diff WLE		age*diff raw		n
		coefficient	p	coefficient	p	
map	Kendall-Tau-b	-0.03	n.s.	-0.04	n.s.	
	Spearman-Rho	-0.04	n.s.	-0.05	n.s.	324
	Pearson	-0.03	n.s.	-0.03	n.s.	
GIS	Kendall-Tau-b	-0.08	0.01	-0.09	0.01	
	Spearman-Rho	-0.10	0.01	-0.11	0.01	591
	Pearson	-0.11	0.01	-0.10	0.01	

Tab. 33: Correlations between age (selection) and difference scores
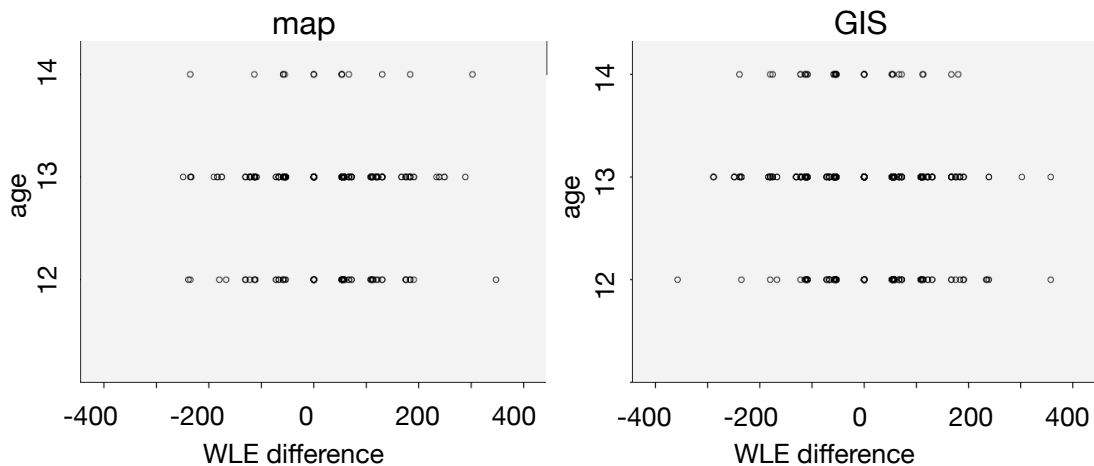
Fig. 40: WLE difference score vs. age (selection) by group

Correlation analyses support that the effect of age on pre-posttest-change observed in Fig. 39 is significant for the GIS group, but the coefficients are small (Fig. 40, Tab. 33).

Two way ANOVAs show only a significant effect for group, with a very low effect size. For the ANOVAs the reduced age was used as fixed effect (Levene's test n.s.) due to small cell sizes and the non-normality of age (Fig. 41, Tab. 34).
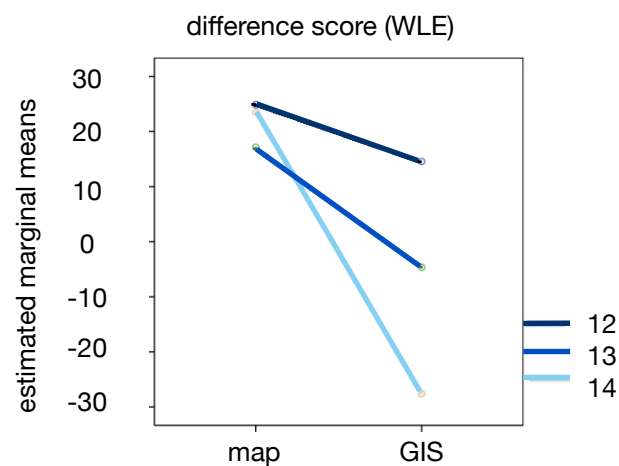
Fig. 41: Two way ANOVA graphical results: interaction: group*age (selection)

	raw				WLE			
	F	**p**	**adjusted R^2**	**partial η^2**	**F**	**p**	**adjusted R^2**	**partial η^2**
group	7.34	0.007		0.008	6.03	0.01		0.01
age	1.73	n.s.	0.01	0.004	1.77	n.s.	0.01	0.004
age*group	0.95	n.s.		0.002	0.83	n.s.		0.002

Tab. 34: Two way ANOVAs: group, age (selection) and interaction

Age as a level 1 predictor in the HLM has no significant effect (Tab. 35).

age	fixed effects		variance components			**explained**
	coefficient	**p**	**u0**	**R**	**p**	
intercept	6.61	n.s.	354.58	10166.58	0.004	0.00
age	-8.47	n.s.				
deviance (parameters)	10506.66 (2)		**homogeneity test**	n.s.	**reliability**	0.42
age and GIS	fixed effects		variance components			**explained**
	coefficient	**p**	**u0**	**R**	**p**	
intercept	19.81	0.002	273.57	10164.39	0.016	0.00
age	-8.47	n.s.				
GIS	-20.78	0.015				0.23
deviance (parameters)	10493.36 (2)		**homogeneity test**	n.s.	**reliability**	0.36

Tab. 35: HLM: age as level 1 predictor models

The tests show mixed results (Tab. 36), thus the hypothesis should be rejected.

		significance	**effect size**
prepost-change within the age sub-groups for the map group	**(t)**	✗	✗
	(W)	✗	
prepost-change within the age sub-groups for the GIS group	**(t)**	✗	✗
	(W)	✗	
difference in prepost-change between the age sub-groups within the map group	**(ANOVA)**	✔	✔
	(KW)	✗ (raw)/ ✔ (WLE)	
difference in prepost-change between the age sub-groups within the GIS group	**(ANOVA)**	✗	✔
	(KW)	✗	
difference in prepost-change between the map and GIS group within the sub-groups	**(t)**	✗	✗
	(MW)	✗	
correlation between prepost-change and age within the map group		✔	✔
correlation between prepost-change and age within the GIS group		✗	✔
two way ANOVA: group, age and interaction		✔	✔
HLM		✔	

Tab. 36: Overview of tests for hypothesis III-a
("On average, there is no effect of age on improvement.")
(✗: contrary to hypothesis, ✔: in line with hypothesis)

7.2.4 Hypothesis III-b: Improvement and sex

III Improvement of achievement: On average ...

b) ... there is no effect of sex on improvement.

Only the female map students' increase is significant (*dz*>0.2). In the GIS group, results are more heterogenous with only the raw values for the female students being significant (*dz*<0.2) (Fig. 42, Tab. 37).

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	male	pre	0.49	0.24	n.s.	< 0.2	n.s.	505.44	107.11	n.s.	< 0.2	n.s.	150
		post	0.52	0.26				515.51	118.48				
	female	pre	0.48	0.22	< 0.001	0.29	< 0.001	504.47	95.39	< 0.001	0.27	0.002	175
		post	0.54	0.25				529.32	115.33				
GIS	male	pre	0.47	0.23	0.07	< 0.2	0.07	497.09	101.62	0.05	< 0.2	0.06	313
		post	0.45	0.25				485.70	115.78				
	female	pre	0.46	0.22	0.03	< 0.2	0.04	495.21	96.44	0.09	< 0.2	0.09	283
		post	0.49	0.25				505.56	112.47				

Tab. 37: Pre- and posttest mean values by group and sex
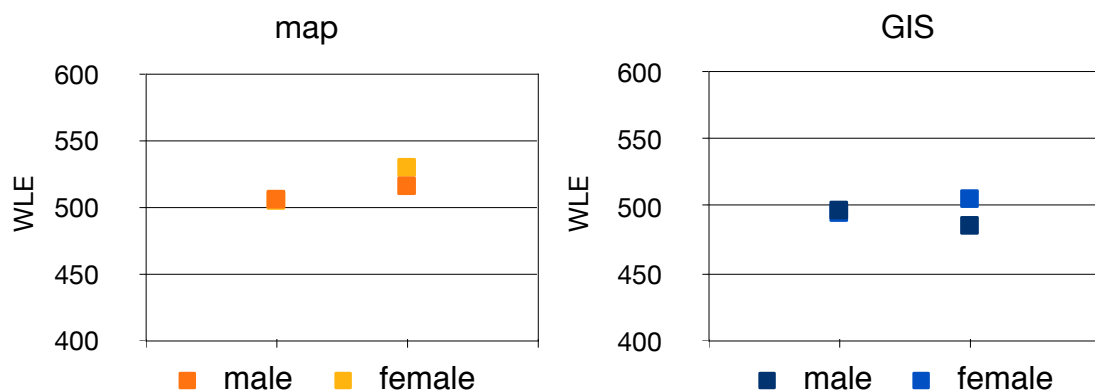
Fig. 42: Pre- and posttest mean values by group and sex

Only in GIS group there is a significant difference between male and female students, in favor of females (*dz*>0.2, Tab. 38). The t-tests have a significant Levene's test in the map (raw *p*=0.02, WLE *p*=0.03) but not in the GIS group. Splitting the sample by sex, there is only a significant difference for male students for the t-test, in favor of the map group (d≅0.2, Fig. 43, Tab. 39; Levene's test raw female *p*=0.05, male *p*=0.095).

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	male	0.03	0.25	n.s.	< 0.2	0.09	10.07	112.36	n.s.	< 0.2	0.095	150
	female	0.06	0.20				24.85	92.05				175
GIS	male	-0.02	0.22	0.005	0.23	0.01	-11.38	102.18	0.009	0.21	0.01	313
	female	0.03	0.22				10.35	100.85				283

Tab. 38: Difference score mean values by group and sex
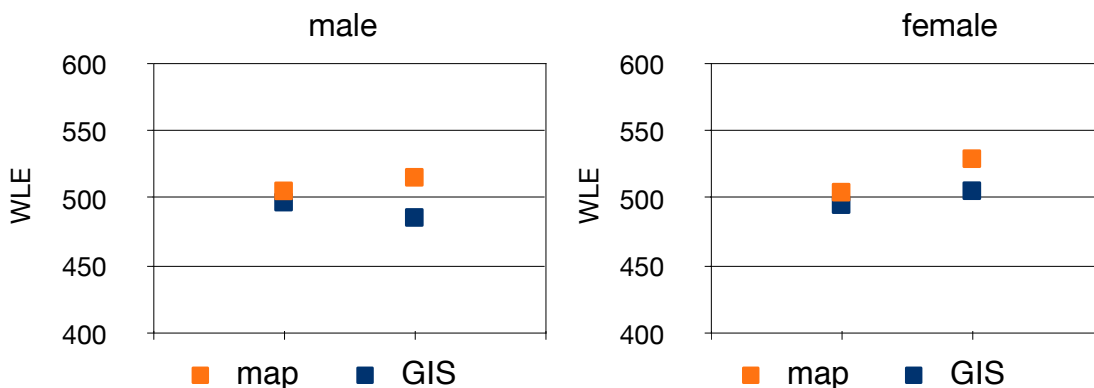
Fig. 43: Pre- and posttest mean values by sex and group

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
male	map	0.03	0.25	0.03	0.21	0.07	10.07	112.36	0.04	0.20	n.s.	150
	GIS	-0.02	0.22				-11.38	102.18				313
female	map	0.06	0.20	n.s.	< 0.2	0.09	24.85	92.05	n.s.	< 0.2	0.097	175
	GIS	0.03	0.22				10.35	100.85				283

Tab. 39: Difference score mean values by sex and group

Two way ANOVAs (Levene's test n.s.) show a significant effect for both group and sex but not the interaction, with very small effect sizes (Fig. 44, Tab. 40, Levene's test raw p=0.08).
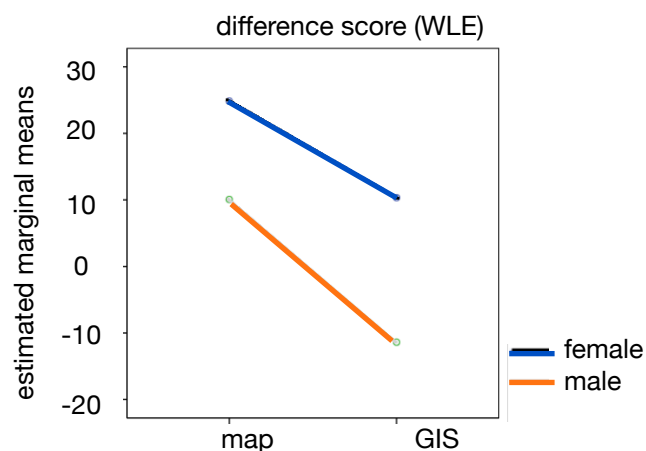
Fig. 44: Two way ANOVA graphical results: interaction group*sex

	raw				WLE			
	F	**p**	**adjusted R²**	**partial η²**	**F**	**p**	**adjusted R²**	**partial η²**
group	6.58	0.01		0.007	6.54	0.01		0.007
sex	7.30	0.007	0.02	0.008	6.75	0.01	0.01	0.007
group*sex	0.45	n.s.		0.00	0.25	n.s.		0.00

Tab. 40: Two way ANOVAs: group, sex and interaction

sex	fixed effects		variance components			explained
	coefficient	**p**	**u0**	**R**	**p**	
intercept	17.02	0.006	316.66	10103.86	0.007	0.01
male	-20.80	0.005				
deviance (parameters)	10499.30 (2)		**homogeneity test**	0.090	**reliability**	0.39
sex and GIS	fixed effects		variance components			explained
	coefficient	**p**	**u0**	**R**	**p**	
intercept	29.13	< 0.001	246.64	10102.02	0.03	0.01
male	-19.49	0.02				
GIS	-20.22	0.006				0.22
deviance (parameters)	10486.47 (2)		**homogeneity test**	0.091	**reliability**	0.33

Tab. 41: HLM: sex as a level 1 predictor model

Using sex as a dummy variable level 1 predictor in the HLM shows a significant negative effect for 'male' (Tab. 41).

The results are mixed (Tab. 42). Consequently, the hypothesis should be rejected.

		significance	effect size
prepost-change within both sexes for the map group	**(t)**	✗	✗
	(W)	✗	
prepost-change within both sexes for the GIS group	**(t)**	✗ (raw)/ ✔ (WLE)	✔
	(W)	✗ (raw)/ ✔ (WLE)	
difference in prepost-change between the two sexes within the map group	**(t)**	✔	✔
	(MW)	✔	
difference in prepost-change between the two sexes within the GIS group	**(t)**	✗	✗
	(MW)	✗	
difference in prepost-change between the map and GIS group within the sub-groups	**(t)**	✗	✗
	(MW)	✔	
two way ANOVA with group, sex and interaction		✗	✔
HLM		✗	

Tab. 42: Overview of tests for hypothesis III-b
        ("On average, there is no effect of sex on improvement.")
        (✗: contrary to hypothesis, ✔: in line with hypothesis)

7.2.5 Hypothesis III-c: Improvement and migration background

III Improvement of achievement: On average, ...

> *c) ... migration background has a negative effect on improvement.*

Only the map without migration background students' improvement is significant ($d>0.2$, Fig. 45, Tab. 43).

There are no significant differences in both groups between students with and without migration background ($d<0.2$, Tab. 44). Only for students without migration background there is a significant difference between map and GIS group ($d>0.2$, Fig. 46, Tab. 45).

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	no mig	pre	0.50	0.23	< 0.001	0.23	< 0.001	509.13	102.85	< 0.001	0.22	0.001	260
		post	0.55	0.25				531.18	113.37				
	mig	pre	0.46	0.22	n.s.	< 0.2	n.s.	492.77	92.15	n.s.	< 0.2	n.s.	65
		post	0.48	0.27				496.34	127.03				
GIS	no mig	pre	0.48	0.22	n.s.	< 0.2	n.s.	498.99	99.23	n.s.	< 0.2	n.s.	510
		post	0.47	0.25				495.78	116.51				
	mig	pre	0.45	0.23	n.s.	< 0.2	n.s.	486.95	101.55	n.s.	< 0.2	n.s.	90
		post	0.46	0.23				493.30	107.55				

Tab. 43: Pre- and posttest mean values by group and migration background
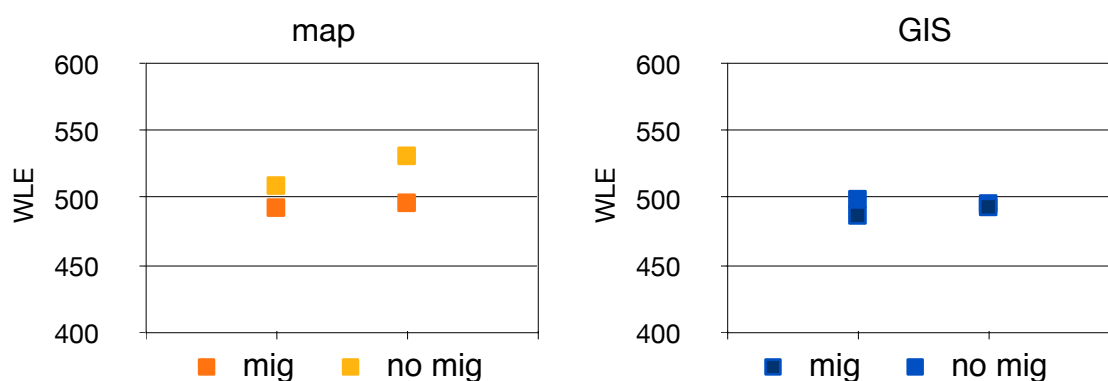
Fig. 45: Pre- and posttest mean values by group and migration background

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	no mig	0.05	0.22	n.s.	< 0.2	n.s.	22.06	98.73	n.s.	< 0.2	n.s.	260
	mig	0.01	0.25				3.56	117.35				65
GIS	no mig	-0.00	0.23	n.s.	< 0.2	n.s.	-3.20	102.41	n.s.	< 0.2	n.s.	510
	mig	0.01	0.22				6.34	98.70				90

Tab. 44: Difference score mean values by group and migration background

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
no mig	**map**	0.05	0.22	0.002	0.24	0.001	22.06	98.73	0.001	0.25	0.002	260
	GIS	-0.00	0.23				-3.20	102.41				510
mig	**map**	0.01	0.25	n.s.	< 0.2	n.s.	3.56	117.35	n.s.	< 0.2	n.s.	65
	GIS	0.01	0.22				6.34	98.70				90

Tab. 45: Difference score mean values by migration background and group
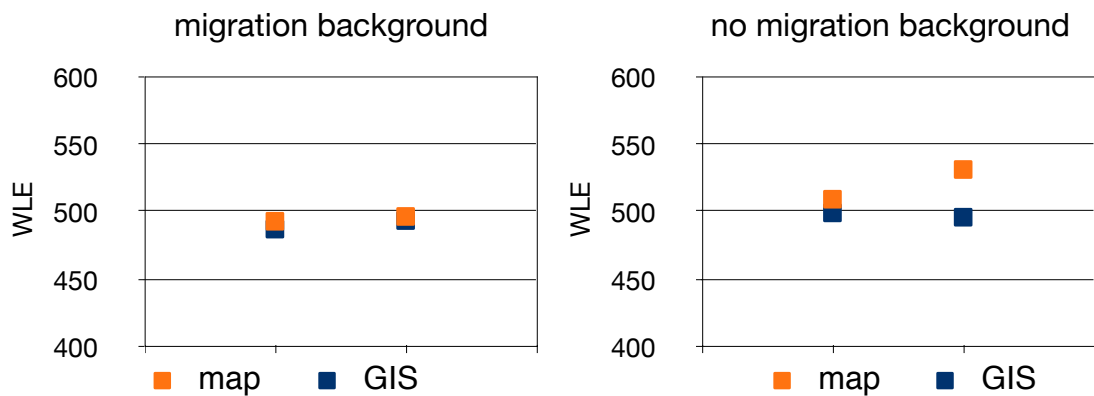
Fig. 46: Pre- and posttest mean values by migration background and group

Two way ANOVAs (Levene's test n.s.) show no significant effect of migration background (Fig. 47, Tab. 46).
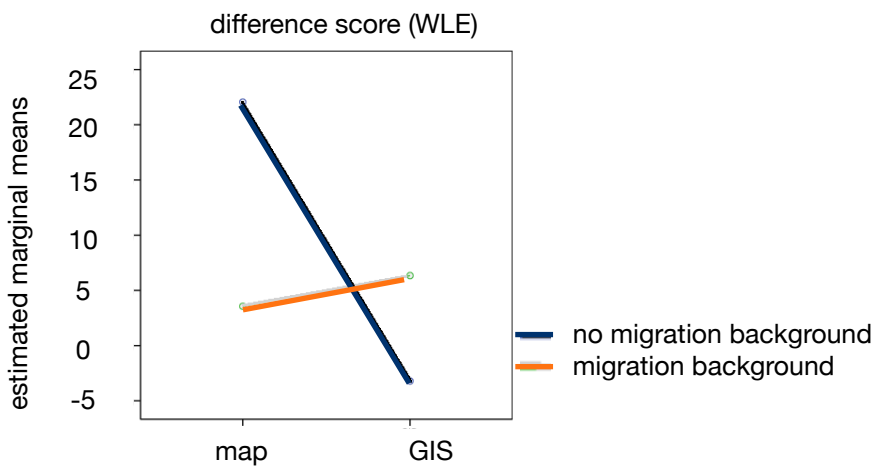
Fig. 47: Two way ANOVA graphical results: interaction group*migration background

	raw				WLE			
	F	p	adjusted R^2	partial η^2	F	p	adjusted R^2	partial η^2
group	1.71	n.s.		0.002	1.50	n.s.		0.002
mig.	0.27	n.s.	0.07	0.00	0.24	n.s.	0.08	0.00
group*mig.	1.86	n.s.		0.002	2.33	n.s.		0.003

Tab. 46: Two way ANOVAs: group, migration background and interaction

migration background	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	6.89	n.s.	358.22	10186.16	0.003	-0.00
migration	-1.82	n.s.				
deviance (parameters)	10507.62 (2)		homogeneity test	n.s.	reliability	0.42
migration background and GIS	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	20.23	0.004	277.20	10183.48	0.02	-0.00
migration	-2.30	n.s.				0.23
GIS	-20.87	0.02				
deviance (parameters)	10494.30 (2)		homogeneity test	n.s.	reliability	0.36

Tab. 47: HLM: migration background as level 1 predictor models

Migration background as a dummy variable in a HLM shows no significant effect (Tab. 47). Discussions with other researchers indicate that consequently, the negative value for variance explained can be ignored and that the higher R value than the intercept model could either be a random occurrence or hint at the variable being negative for the model. SSI Central (n.d.-c, p. 137, with reference to Raudenbush & Bryk 2002, pp. 149-152) states that "[...] if a truly nonsignificant variable enters the model, it is mathematically possible under maximum likelihood to observe a slight increase in the residual variance [...]".

The results are mixed (Tab. 48). Thus, the hypothesis should be rejected.

		significance	effect size
prepost-change within both sub-groups for the map group	(t)	✔	✔
	(W)	✔	
prepost-change within both sub-groups for the GIS group	(t)	✘	✘
	(W)	✘	
difference in prepost-change between the two sub-groups within the map group	(t)	✘	✘
	(MW)	✘	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✘	✘
	(MW)	✘	
difference in prepost-change between the map and GIS group within the sub-groups	(t)	✘	✘
	(MW)	✘	
two way ANOVA: group, migration background and interaction		✘	✘
HLM		✘	

Tab. 48: Overview of tests of hypothesis III-c
("On average, migration background has a negative effect on improvement.")
(✘: contrary to hypothesis, ✔: in line with hypothesis)

7.2.6 Hypothesis III-d: Improvement and language background

III Improvement of achievement: On average …

d) there is no effect of language background on improvement.

Cell sizes for students with a total language count of five or six are too small, leaving 920 students. Only the map students' speaking/learning three languages increase is significant (*dz*>0.2). The four language students' decrease also has an effect size above 0.2 (Fig. 48, Tab. 49). In both groups, two language students have the lowest pretest score.
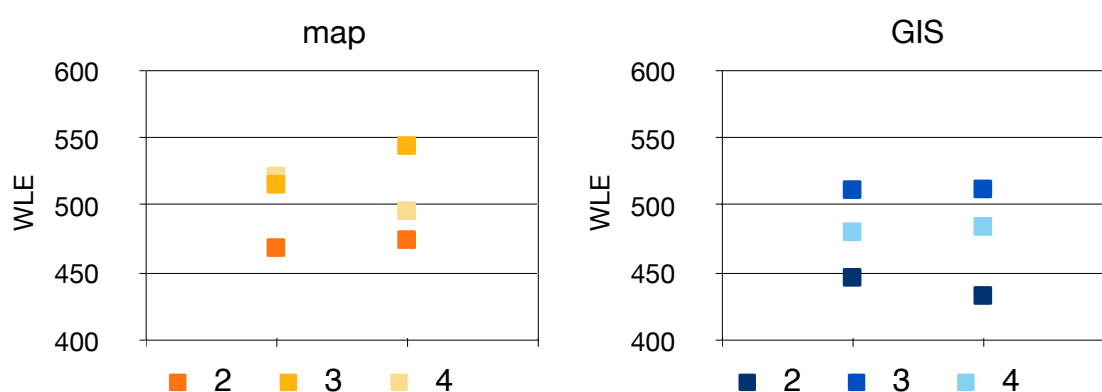
Fig. 48: Pre- and posttest mean values by group and total language (selection)

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	2	pre	0.40	0.20	n.s.	< 0.2	n.s.	468.61	88.34	n.s.	< 0.2	n.s.	63
		post	0.42	0.23				474.53	100.75				
	3	pre	0.51	0.23	< 0.001	0.29	< 0.001	515.60	103.31	< 0.001	0.27	< 0.001	215
		post	0.58	0.24				544.32	111.54				
	4	pre	0.53	0.21	n.s.	0.21	n.s.	521.68	85.46	n.s.	0.23	n.s.	45
		post	0.48	0.28				496.07	131.43				
GIS	2	pre	0.36	0.20	n.s.	< 0.2	n.s.	446.53	91.18	n.s.	< 0.2	0.09	105
		post	0.34	0.25				433.06	120.65				
	3	pre	0.50	0.22	n.s.	< 0.2	n.s.	511.62	98.05	n.s.	< 0.2	n.s.	429
		post	0.51	0.24				512.15	110.53				
	4	pre	0.43	0.23	n.s.	< 0.2	n.s.	480.36	97.39	n.s.	< 0.2	n.s.	63
		post	0.44	0.23				484.32	103.00				

Tab. 49: Pre- and posttest mean values by group and total languages (selection)

In the ANOVAs, only in the map group there is a significant effect for total number of languages, with an effect size below the threshold (Levene's test raw *p*=0.01, WLE *p*=0.03, Tab. 50, Fig. 49).

| | | diff raw | | ANOVA raw | | | | KW | diff WLE | | ANOVA WLE | | | | KW | |
		M	SD	F	p	adj. R²	par-tial η²	p	M	SD	F	p	adj. R²	par-tial η²	p	n
map	2	0.02	0.18						5.92	78.83						63
	3	0.07	0.23	5.56	0.004	0.03	0.03	0.004	28.72	104.63	5.84	0.003	0.03	0.04	0.003	215
	4	-0.05	0.24						-25.61	110.15						45
GIS	2	-0.02	0.22						-13.47	101.42						105
	3	0.00	0.23	0.54	n.s.	-0.00	0.00	n.s.	0.53	103.73	0.90	n.s.	0.00	0.00	n.s.	429
	4	0.00	0.20						3.95	88.92						63

Tab. 50: Difference score mean values by group and total languages (selection)
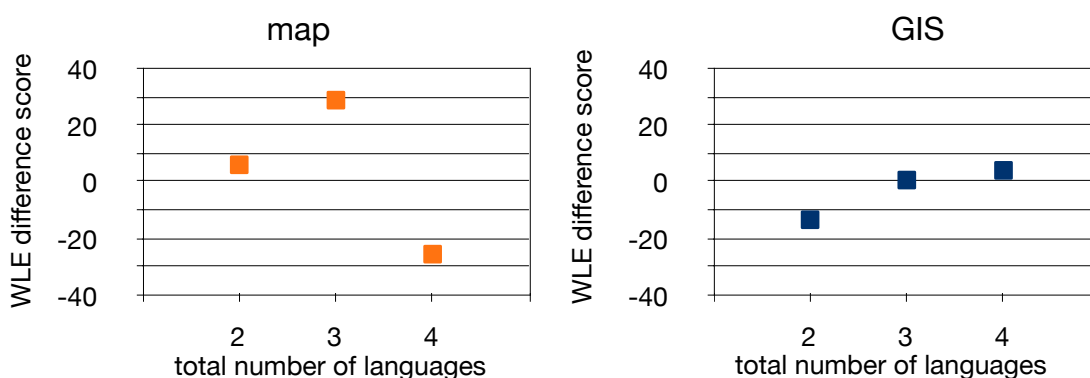
Fig. 49: WLE difference score group and total number of languages (selection)

Splitting the sample by total languages, the Levene's test for two language students is significant for WLE ($p=0.02$; raw $p=0.08$). There are significant differences between map and GIS group for three language-students, in favor of map students ($dz>0.2$). Most effect sizes are above 0.2 (Tab. 51, Fig. 50; Levene's test four language students WLE $p=0.05$, raw $p=0.07$).
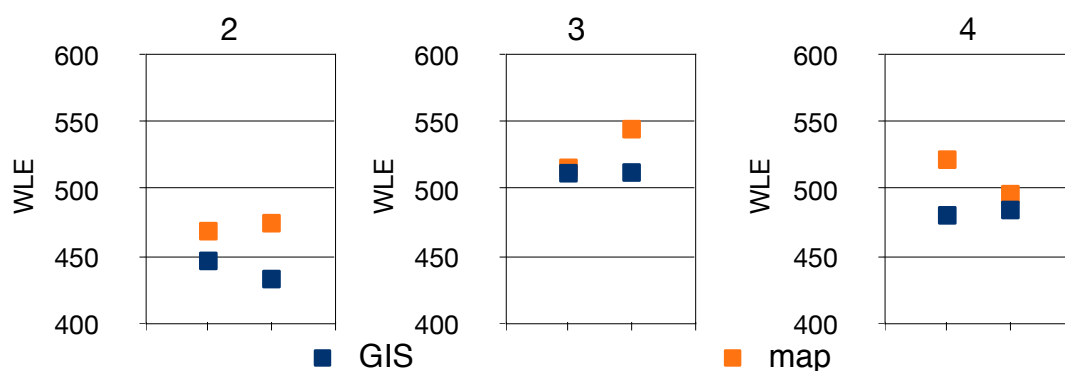
Fig. 50: Pre- and posttest mean values by total languages (selection) and group

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
2	map	0.02	0.18	n.s.	< 0.2	n.s.	5.92	78.83	n.s.	0.21	n.s.	63
	GIS	-0.02	0.22				-13.47	101.42				105
3	map	0.07	0.23	0.001	0.27	0.001	28.72	104.63	0.001	0.27	0.001	215
	GIS	0.00	0.23				0.53	103.73				429
4	map	-0.05	0.24	n.s.	0.25	n.s.	-25.61	110.15	n.s.	0.30	n.s.	45
	GIS	0.00	0.20				3.95	88.92				63

Tab. 51: Difference score mean values by total languages (selection) and group

Two way ANOVAs (Levene's test WLE p=0.09) show significant effects for total language and interaction, with effect sizes below the threshold (Fig. 51, Tab. 52). Correlations are not significant and have small coefficients (Fig. 52, Tab. 53).
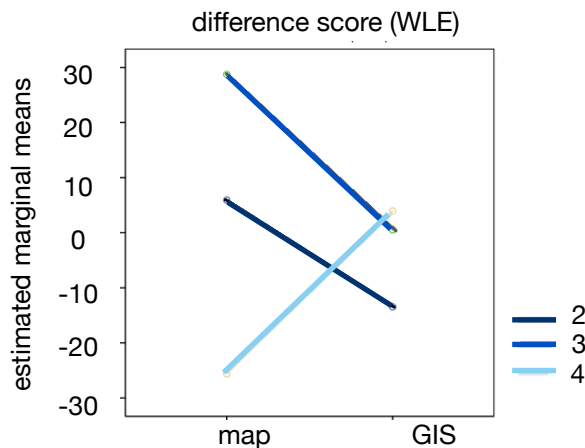
Fig. 51: Two way ANOVA graphical results: interaction total languages (selection)*group

	raw				WLE			
	F	p	adjusted R^2	partial η^2	F	p	adjusted R^2	partial η^2
group	0.50	n.s.	0.02	0.001	0.45	n.s.	0.02	0.000
tot. lan.	4.07	0.02		0.009	4.08	0.02		0.009
group*tot.lan.	3.03	0.05		0.007	3.59	0.03		0.008

Tab. 52: Two way ANOVAs: group, total languages (selection) and interaction

		total language*diff WLE		total language*diff raw		n
		coefficient	p	coefficient	p	
map	Kendall-Tau-b	-0.03	n.s.	-0.04	n.s.	323
	Spearman-Rho	-0.04	n.s.	-0.05	n.s.	
	Pearson	-0.06	n.s.	-0.06	n.s.	
GIS	Kendall-Tau-b	0.05	n.s.	0.04	n.s.	597
	Spearman-Rho	0.06	n.s.	0.05	n.s.	
	Pearson	0.05	n.s.	0.04	n.s.	

Tab. 53: Correlations between total language (selection) and difference scores
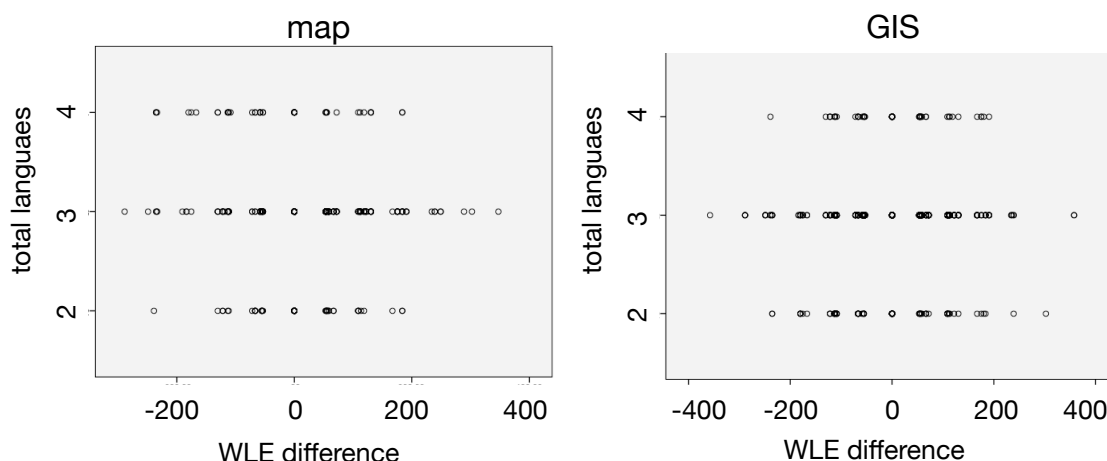
Fig. 52: WLE difference score vs. total languages (selection) by group

Excluding the eleven students speaking "1 other" as family language leaves 921 students. Only map students speaking only German at home increase significantly (*dz*>0.2, Fig. 53, Tab. 54).
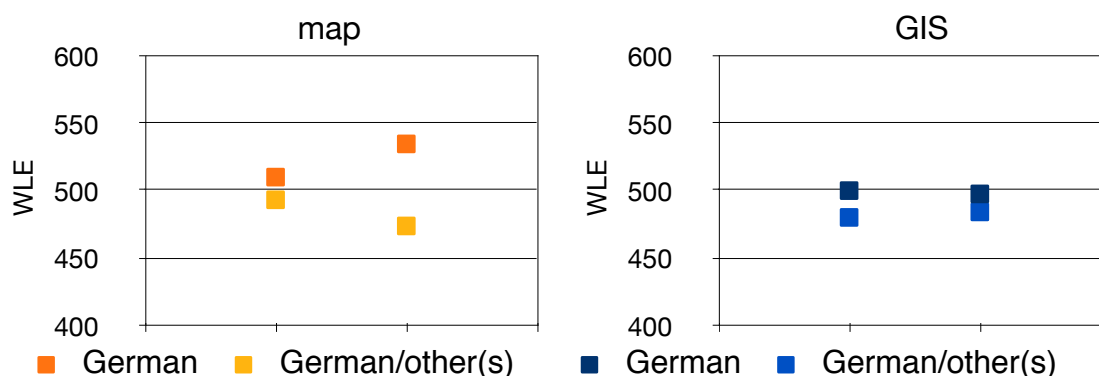
Fig. 53: Pre- and posttest mean values by group and family languages (selection)

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	German	pre	0.50	0.23	< 0.001	0.26	< 0.001	509.82	101.67	< 0.001	0.25	< 0.001	271
		post	0.56	0.24				534.47	111.69				
	German/other(s)	pre	0.46	0.22	n.s.	< 0.2	n.s.	493.00	91.53	n.s.	< 0.2	n.s.	51
		post	0.43	0.28				473.57	135.26				
GIS	German	pre	0.48	0.23	n.s.	< 0.2	n.s.	499.80	100.69	n.s.	< 0.2	n.s.	521
		post	0.48	0.25				497.36	116.16				
	German/other(s)	pre	0.43	0.21	n.s.	< 0.2	n.s.	479.91	91.26	n.s.	< 0.2	n.s.	78
		post	0.44	0.24				483.99	109.31				

Tab. 54: Pre- and posttest mean values by group and family languages (selection)

In the map group there are significant differences by family language background, in favor "German only" students (d>0.2, Tab. 55, Levene's test WLE map *p*=0.07). Within a family language group there is a significant difference for

the "German only" subgroup, in favor of map students ($d>0.2$). However, also the effect size for the WLE score in the "German/other(s)" subgroup is above 0.2, in favor of GIS students (Tab. 56, Fig. 54).

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	German	0.06	0.22	0.01	0.37	0.03	24.66	98.69	0.005	0.41	0.02	271
	German/other(s)	-0.03	0.25				-19.43	115.29				51
GIS	German	-0.00	0.22	n.s.	< 0.2	n.s.	-2.44	101.81	n.s.	< 0.2	n.s.	521
	German/other(s)	0.01	0.23				4.07	103.23				78

Tab. 55: Difference score mean values by group and family languages (selection)
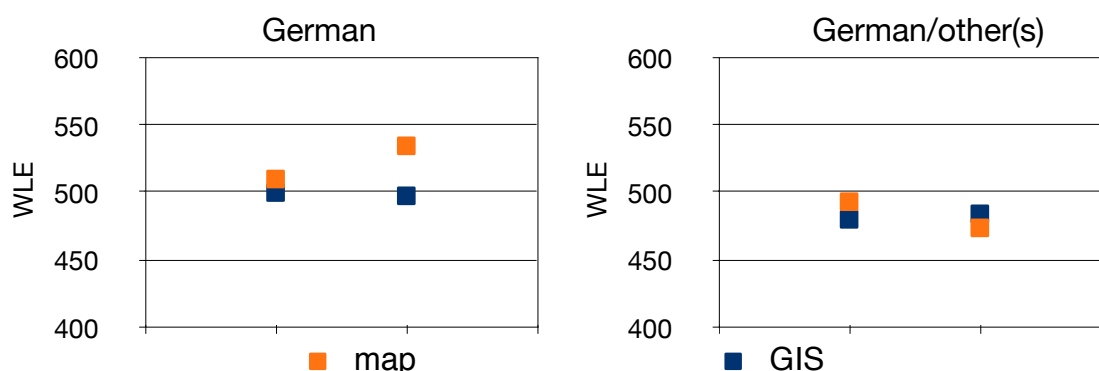
Fig. 54: Pre- and posttest mean values by family languages (selection) and group

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
German	map	0.06	0.22	0.001	0.26	0.001	24.66	98.69	< 0.001	0.27	0.001	271
	GIS	-0.00	0.22				-2.44	101.81				521
German/ other(s)	map	-0.03	0.25	n.s.	< 0.2	n.s.	-19.43	115.29	n.s.	0.21	n.s.	51
	GIS	0.01	0.23				4.07	103.23				78

Tab. 56: Difference score mean values by family languages (selection) and group

Two way ANOVAs (Levene's test n.s.) show a significant effect only for the interaction (Fig. 55, Tab. 57).

	raw				WLE			
	F	p	adjusted R²	partial η²	F	p	adjusted R²	partial η²
group	0.16	n.s.	0.01	0.00	0.03	n.s.	0.01	0.00
fam. lan.	3.10	0.08		0.00	3.58	0.06		0.00
group*fam. lan.	4.98	0.03		0.01	6.49	0.01		0.01

Tab. 57: Two way ANOVAs: group, family languages (selection) and interaction
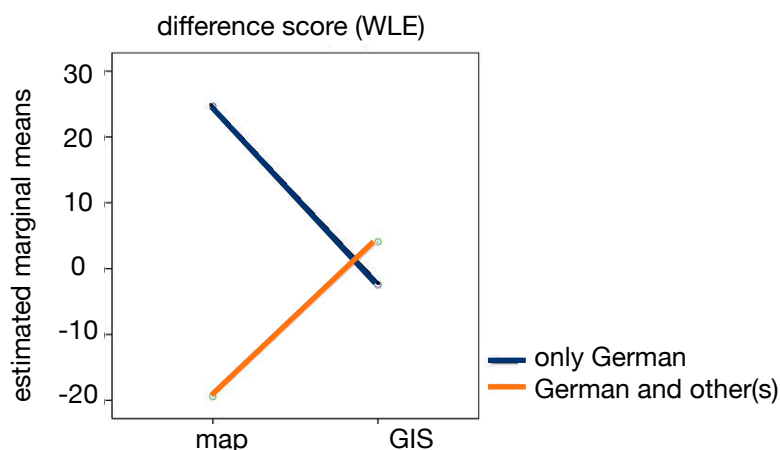
Fig. 55: Two way ANOVA graphical results: interaction family languages (selection)*group

HLMs show no significant effect for either language background variable (Tab. 58) in individual analyses.

language background	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	6.61	n.s.	355.03	10156.10	0.003	0.00
total no. language	-11.75	0.09				
deviance (parameters)	10505.53 (2)		homogeneity test	n.s.	reliability	0.42
intercept	8.68	0.07	390.01	10141.66	0.002	0.00
fam. lang. other	-15.44	n.s.				
deviance (parameters)	10505.26 (2)		homogeneity test	n.s.	reliability	0.44
language background and GIS	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	18.81	0.002	274.04	10153.91	0.02	0.00
total no. language	-11.75	0.09				
GIS	-20.78	0.02				0.23
deviance (parameters)	10492.23 (2)		homogeneity test	n.s.	reliability	0.36
intercept	22.15	0.002	305.29	10139.63	0.009	0.00
fam. lang. other	-15.54	n.s.				
GIS	-21.20	0.02				0.22
deviance (parameters)	10491.92 (2)		homogeneity test	n.s.	reliability	0.38

Tab. 58: HLM: language background variables as level 1 predictor models
(individual analyses)

The results are mixed (Tab. 59). Consequently, the hypothesis should be rejected.

total number of languages		significance	effect size
prepost-change within the sub-groups for the map group	(t)	✗	✗
	(W)	✗	
prepost-change within the sub-groups for the GIS group	(t)	✔	✔
	(W)	✔	
difference in prepost-change between the sub-groups within the map group	(ANOVA)	✗	✔
	(KW)	✗	
difference in prepost-change between the sub-groups within the GIS group	(ANOVA)	✔	✔
	(KW)	✔	
difference in prepost-change between the map and GIS group within the sub-groups	(t)	✗	✗
	(MW)	✗	
correlation between total language and systemic thinking difference score for the map group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
correlation between total language and systemic thinking difference score for the GIS group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
two way ANOVA with group, total number of languages and interaction		✗	✔
HLM		✔	
family language background		significance	effect size
prepost-change within both sub-groups for the map group	(t)	✗	✗
	(W)	✗	
prepost-change within both sub-groups for the GIS group	(t)	✔	✔
	(W)	✔	
difference in prepost-change between the two sub-groups within the map group	(t)	✗	✗
	(MW)	✗	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the map and GIS group within the sub-groups	(t)	✗	✗
	(MW)	✗	
two way ANOVA: group, family language background and interaction		✔	✔
HLM		✔	

Tab. 59: Overview of tests for hypothesis III-d
("On average, there is no effect of language background on improvement.")
(✗: contrary to hypothesis, ✔: in line with hypothesis)

7.2.7 Hypothesis III-e: Improvement and pre-achievement

III Improvement of achievement: On average, ...

e) ... pre-achievement has an effect on improvement.

This hypothesis can be divided into two parts: by overall higher achievement (stream) and concrete higher achievement (pretest score). The pretest score can be examined both by using the complete range and by a binary classification.

Only the map high stream students' improvement is significant ($dz > 0.2$, Fig. 56, Tab. 60). Fig. 56 also highlights the lower pre-score for mid stream students.

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	mid	pre	0.39	0.19	n.s.	< 0.2	n.s.	465.32	80.21	n.s.	< 0.2	n.s.	107
		post	0.41	0.22				468.63	98.43				
	high	pre	0.54	0.23	0.001	0.23	0.001	524.90	104.11	0.001	0.22	0.002	222
		post	0.59	0.25				549.42	116.89				
GIS	mid	pre	0.37	0.20	n.s.	< 0.2	0.098	453.31	87.71	0.05	< 0.2	0.07	203
		post	0.35	0.23				439.42	112.26				
	high	pre	0.52	0.22	n.s.	< 0.2	n.s.	519.16	97.89	n.s.	< 0.2	n.s.	400
		post	0.53	0.23				523.86	105.59				

Tab. 60: Pre- and posttest mean values by group and stream
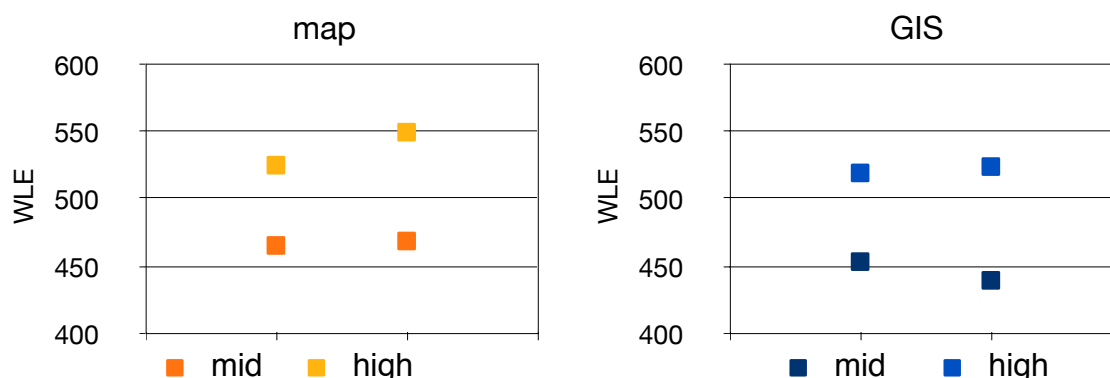
Fig. 56: Pre- and posttest mean values by group and stream

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	mid	0.02	0.20	n.s.	< 0.2	0.095	3.31	85.79	0.06	0.22	0.06	107
	high	0.05	0.24				24.53	109.44				222
GIS	mid	-0.02	0.21	0.05	< 0.2	0.03	-13.89	100.95	0.03	< 0.2	0.03	203
	high	0.01	0.23				4.70	101.56				400

Tab. 61: Difference score mean values by group and stream

Within the map group Levene's test is significant (raw $p=0.02$, WLE $p=0.003$). Only the difference between mid and high stream students in the GIS WLE sub-group is significant. However, the effect size for the GIS group (WLE) is below 0.2, while that for the map group is above 0.2 (Tab. 61). Within one stream Levene's test is not

significant except for the the mid stream WLE value (*p*=0.03, high stream WLE *p*=0.08). Only the high stream differences are significant (*d*<0.2, Fig. 57, Tab. 62).

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
mid	map	0.02	0.20	n.s.	< 0.2	0.06	3.31	85.79	n.s.	< 0.2	n.s.	222
	GIS	-0.02	0.21				-13.89	100.95				400
high	map	0.05	0.24	0.03	< 0.2	0.02	24.53	109.44	0.02	< 0.2	0.02	107
	GIS	0.01	0.23				4.70	101.56				203

Tab. 62: Difference score mean values by stream and group
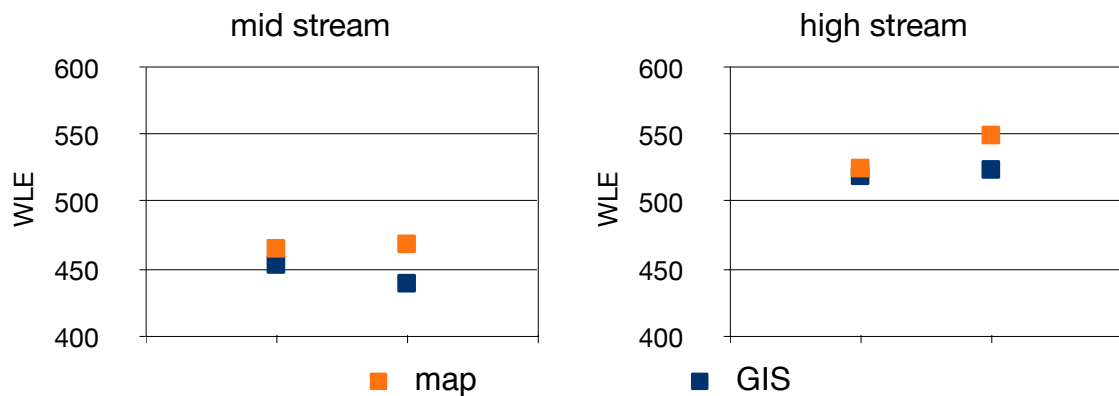
Fig. 57: Pre- and posttest mean values by stream and group

Two way ANOVAs (Levene's test n.s. for raw, *p*=0.03 for WLE) show a significant effect only for group and stream, but not the interaction (Fig. 58, Tab. 63).
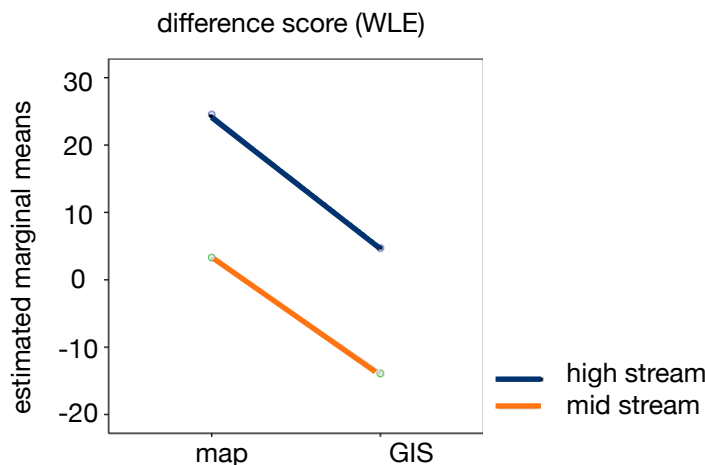
Fig. 58: Two way ANOVA graphical results: interaction group*stream

	raw				WLE			
	F	p	adjusted R^2	partial η^2	F	p	adjusted R^2	partial η^2
group	6.39	0.01		0.01	6.23	0.01		0.01
stream	5.45	0.02	0.01	0.01	7.20	0.007	0.01	0.01
group*stream	0.00	n.s.		0.00	0.03	n.s.		0.00

Tab. 63: Two way ANOVAs: group, stream and interaction

Correlations between pre-score and difference score show a significant negative relationship, i.e. the higher the difference, the lower the pretest score (Fig. 59, Tab. 64). This could be due to students with a low pretest score not having that much room to decrease, and students scoring well on the pretest not being able to improve that much. There are twelve (GIS) and seven (map) students who score 1 on the pretest as well as 18 (GIS) and six (map) students who score 0.

		pre WLE*diff WLE		pre raw*diff raw		n
		coefficient	p	coefficient	p	
map	Kendall-Tau-b	-0.22	< 0.001	-0.26	< 0.001	329
	Spearman-Rho	-0.29	< 0.001	-0.34	< 0.001	
	Pearson	-0.34	< 0.001	-0.38	< 0.001	
GIS	Kendall-Tau-b	-0.24	< 0.001	-0.28	< 0.001	603
	Spearman-Rho	-0.31	< 0.001	-0.36	< 0.001	
	Pearson	-0.35	< 0.001	-0.38	< 0.001	

Tab. 64: Correlations between pre-scores and difference scores
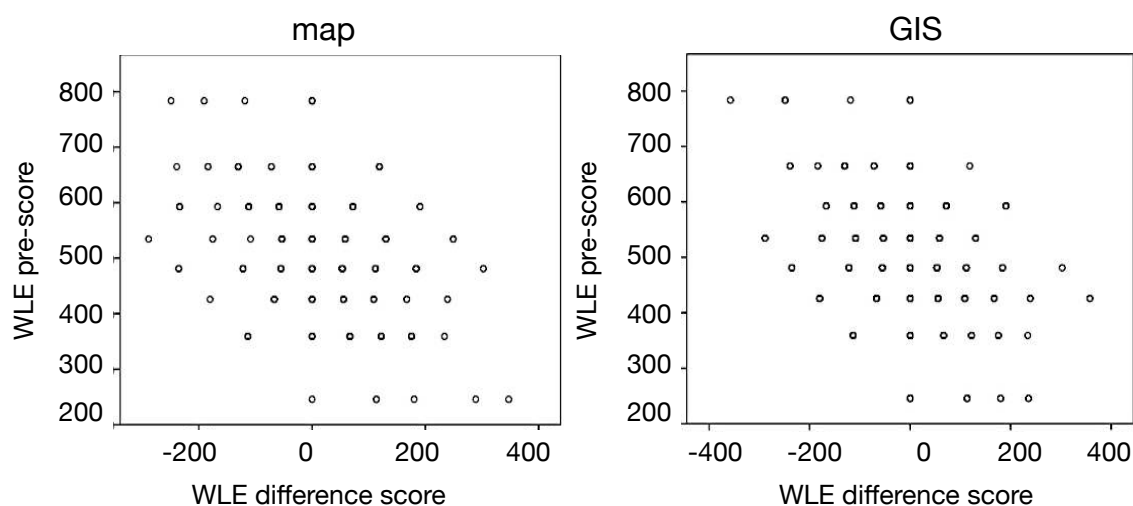
Fig. 59: Pre-score by difference score

ANCOVAs (Levene's test n.s.) with pre-score as co-variate show significant effects for both group and pre-score (Tab. 65), with the effect size for pre-score being larger than for group.

	raw				WLE			
	F	p	adjusted R^2	partial η^2	F	p	adjusted R^2	partial η^2
pre-score	153.48	< 0.001	0.15	0.14	123.69	< 0.001	0.12	0.12
group	11.77	0.001		0.01	11.36	0.001		0.01

Tab. 65: ANCOVAs: pre-score and group

In the map group, only those scoring below 500 in the pretest improve significantly (*dz*>0.2). In the GIS group, the students scoring below 500 in the pretest improve significantly, while those scoring more than 500 decrease significantly (both *dz*>0.2, Fig. 60, Tab. 66).

WLE			M	SD	p	|dz|	p (W)	n
map	< 500	pre	432.85	58.54	< 0.001	0.39	< 0.001	181
		post	471.82	103.60				
	> 500	pre	594.39	63.45	n.s.	< 0.2	n.s.	148
		post	585.91	101.90				
GIS	< 500	pre	427.62	61.31	< 0.001	0.22	0.002	336
		post	450.19	106.48				
	> 500	pre	584.28	62.85	< 0.001	0.34	< 0.001	267
		post	552.37	98.92				

Tab. 66: Pre- and posttest mean values by groups by pre-score (bin.)
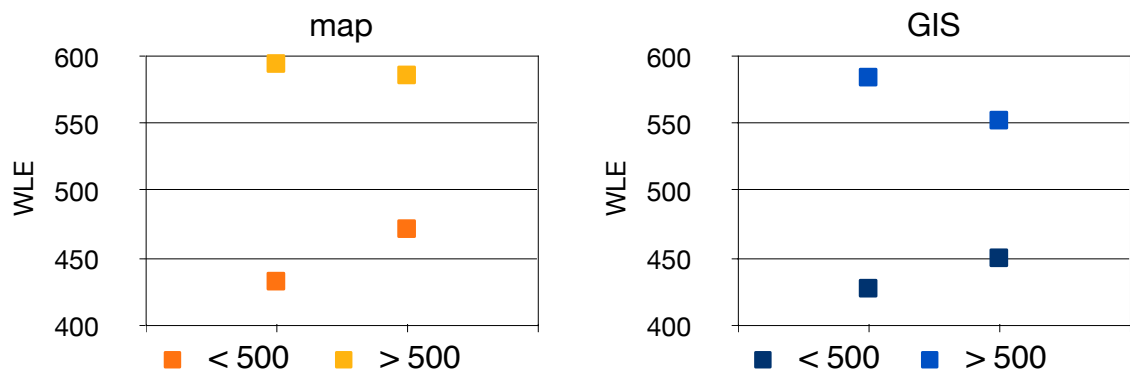
Fig. 60: Pre- and posttest mean values by group and pre-score (bin.)

diff WLE		M	SD	p	|d|	p (MW)	n
map	< 500	38.97	100.07	< 0.001	0.47	< 0.001	181
	> 500	-8.48	100.14				148
GIS	< 500	22.57	101.45	< 0.001	0.56	< 0.001	336
	> 500	-31.91	93.62				267

Tab. 67: Difference score mean values by group and pre-score (bin.)

In both groups the differences between the below 500 and above 500 pre-score students are significant with effect sizes above 0.2 (Tab. 67). Students scoring below 500 in the pretest show no significant difference between map and GIS, while in the above 500 group, the difference is significant with map students decreasing less (*d*>0.2, Fig. 61, Tab. 68).
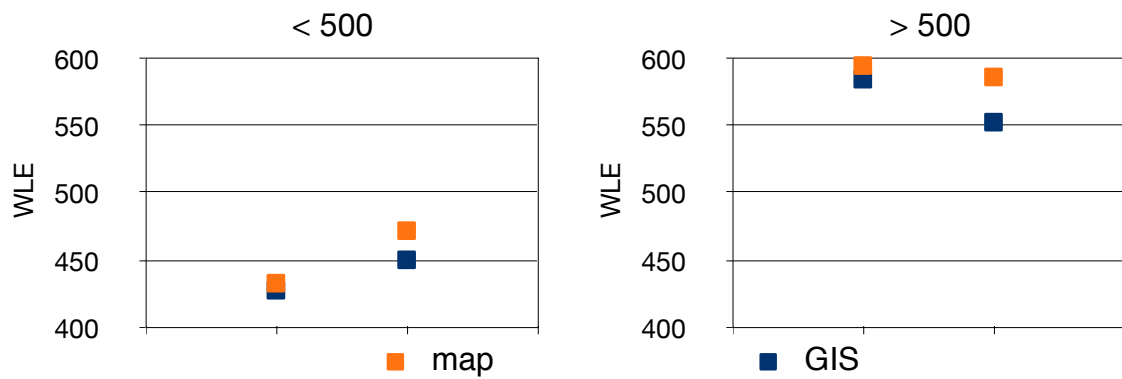
Fig. 61: Pre- and posttest mean values by pre-score (bin.) and group

diff WLE		M	SD	p	|d|	p (MW)	n
< 500	**map**	38.97	100.07	0.08	< 0.2	n.s.	181
	GIS	22.57	101.45				336
> 500	**map**	-8.48	100.14	0.02	0.24	0.02	148
	GIS	-31.91	93.62				267

Tab. 68: Difference score mean values by pre-score (bin.) and group

A two way ANOVA (Levene's test n.s.) shows significant effects for group and pre-score but not for the interaction (Fig. 62, Tab. 69).
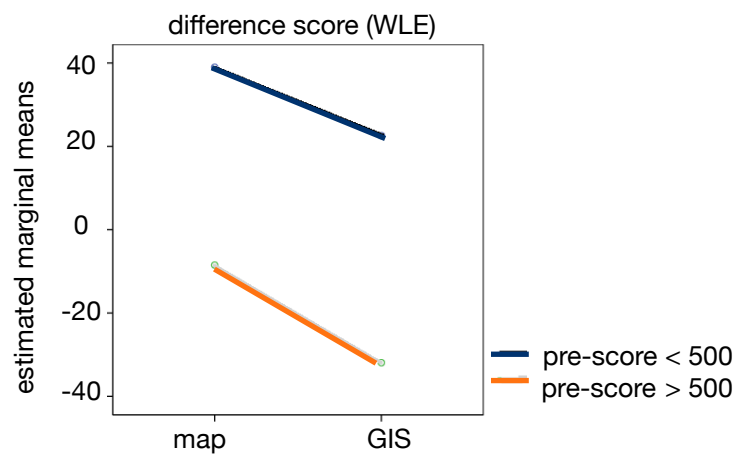
Fig. 62: Two way ANOVA graphical results: interaction group*pre-score (bin.)

WLE	F	p	adjusted R²	partial η²
group	8.56	0.004		0.01
pre-score bin.	56.03	< 0.001	0.07	0.06
group*pre-score bin.	0.27	n.s.		0.000

Tab. 69: Two way ANOVA: group, pre-score (bin.) and interaction

stream	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	-6.50	n.s.	278.82	10179.61	0.014	0.21
high stream	19.82	0.02				
deviance (parameters)	10503.18 (2)		homogeneity test	n.s.	reliability	0.36
stream and GIS	**fixed effects**		**variance components**			**explained**
	coefficient	p	u0	R	p	
intercept	6.76	n.s.	200.06	10178.29	0.054	0.43
high stream	19.09	0.01				
GIS	-20.03	0.01				
deviance (parameters)	10489.87 (2)		homogeneity test	n.s.	reliability	0.29

Tab. 70: HLM: stream as level 2 predictor models

pre-achievement	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	6.55	n.s.	428.24	8517.38	< 0.001	0.16
pre-score	-0.44	< 0.001				
deviance (parameters)	10370.15 (2)		homogeneity test	**0.015**	reliability	0.51
intercept	31.95	< 0.001	614.07	9318.47	< 0.001	0.08
pre-score > 500	-57.11	< 0.001				
deviance (parameters)	10443.66 (2)		homogeneity test	n.s.	reliability	0.57
pre-achievement and GIS	**fixed effects**		**variance components**			**explained**
	coefficient	p	u0	R	p	
intercept	19.72	0.002	348.19	8515.22	0.001	0.16
pre-score	-0.44	< 0.001				0.19
GIS	-20.79	0.02				
deviance (parameters)	10356.82 (2)		homogeneity test	**0.015**	reliability	0.45
intercept	45.22	< 0.001	530.27	9318.30	< 0.001	0.08
pre-score > 500	-56.73	< 0.001				0.14
GIS	-21.29	0.03				
deviance (parameters)	10431.02 (2)		homogeneity test	n.s.	reliability	0.53

Tab. 71: HLM: pre-achievement variables as level 1 predictors models

Adding stream as a level 2 predictor shows a significant positive effect for high stream students. 21.2% of the class variation can be explained by stream, with stream and group together explaining 43.5% (Tab. 70). Using the WLE pre-score as level 1 predictor shows a significant negative effect, explaining 16.3% of the level 1 variance. Using the binary pre-score, the model explains 8.4% of the level 1 variance, with a significant negative effect for high pre-score. Com-

bining (binary) pre-score and GIS, 16.3% (8.4%) of level 1 and 18.7% (13.6%) of level 2 variance is explained (Tab. 71).

The results are mostly, but not all, in favor of the hypothesis (Tab. 72).

stream		significance	effect size
prepost-change within the sub-groups for the map group	(t)	✔	✔
	(W)	✔	
prepost-change within the sub-groups for the GIS group	(t)	✘	✘
	(W)	✘	
difference in prepost-change between the two sub-groups within the map group	(t)	✘	✘ (raw)/ ✔ (WLE)
	(MW)	✘	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✘ (raw)/ ✔ (WLE)	✘
	(MW)	✔	
difference in prepost-change between the map and GIS group within each stream	(t)	✔	✘
	(MW)	✔	
two way ANOVA: group, stream and interaction		✔	✘
HLM		✔	
pre-score		significance	effect size
correlation between pre-score and difference score for the map group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
correlation between pre-score and difference score for the GIS group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
ANCOVA: pre-score and group		✔	✔
HLM		✔	
pre-score (bin.)		significance	effect size
prepost-change within the sub-groups for the map group	(t)	✔	✔
	(W)	✔	
prepost-change within the sub-groups for the GIS group	(t)	✘	✘
	(W)	✘	
difference in prepost-change between the two sub-groups within the map group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the map and GIS group within the pre-score sub-groups	(t)	✔	✔
	(MW)	✔	
two way ANOVA: group, pre-score (bin.) and interaction		✔	✔
HLM		✔	

Tab. 72: Overview of tests for hypothesis III-e
("On average, pre-achievement has an effect on improvement.")
(✘: contrary to hypothesis, ✔: in line with hypothesis)

7.2.8 Hypothesis III-f: Improvement and pre-spatial thinking

III Improvement of achievement: On average ....

f) ... there is no effect of pre-spatial thinking score on improvement.

The correlation between pre-spatial thinking score and systemic thinking difference score is not significant (Fig. 63, Tab. 73). The ANCOVAs (Levene's test n.s.) show significant effects for pre-spatial thinking skills and group in the WLE, and only group in the raw score analysis (Tab. 74). However, effect sizes are very small.

		pre-WLE*diff WLE		pre raw*raw		n
		coefficient	p	coefficient	p	
map	**Kendall-Tau-b**	0.07	n.s.	0.005	n.s.	
	Spearman-Rho	0.09	n.s.	0.005	n.s.	329
	Pearson	0.09	n.s.	0.04	n.s.	
GIS	**Kendall-Tau-b**	0.05	0.09	0.03	n.s.	
	Spearman-Rho	0.07	0.09	0.05	n.s.	603
	Pearson	0.07	n.s.	0.04	n.s.	

Tab. 73: Correlations pre-spatial thinking score and systemic thinking difference score by group
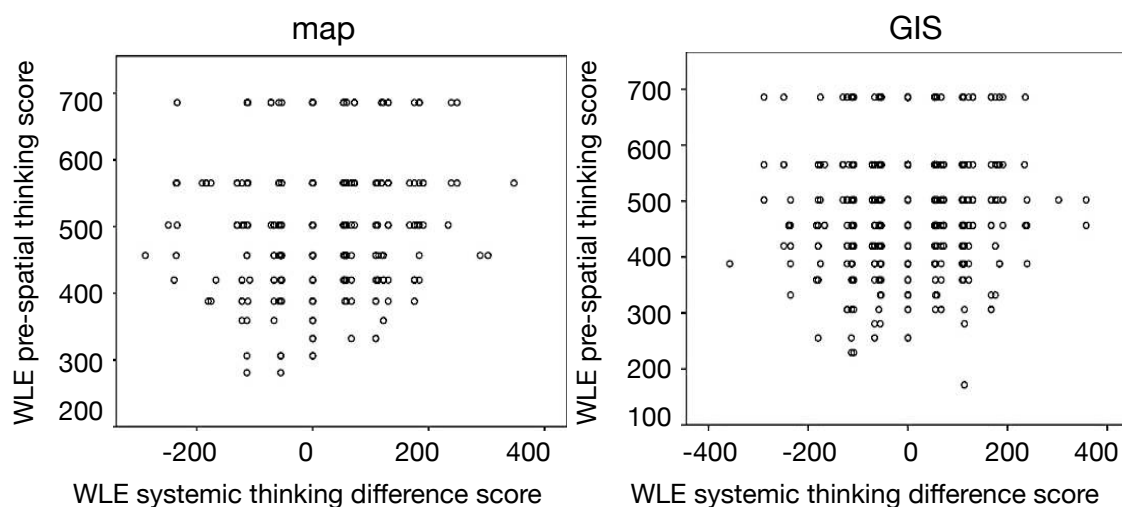
Fig. 63: Pre-spatial thinking score by systemic thinking difference score

	raw				WLE			
	F	**p**	**adjusted R²**	**partial η²**	**F**	**p**	**adjusted R²**	**partial η²**
pre-spatial-thinking-score	1.19	n.s.	0.01	0.00	5.32	0.02	0.01	0.01
group	6.91	0.009		0.01	6.48	0.01		0.01

Tab. 74: ANCOVA: pre-spatial thinking score and group

Only map high scoring students improve significantly ($d>0.2$) (Tab. 75). There are no significant differences between high and low scoring students in both

groups ($d<0.2$, Tab. 76, Fig. 64). Splitting the sample by pre-spatial thinking score, there are no significant differences between map and GIS students ($d<0.2$, Fig. 65, Tab. 77). A two way ANOVA (Levene's test n.s.) shows a significant effect only for group, with very small effect sizes (Fig. 66, Tab. 78).

WLE			M	SD	p	|dz|	p (W)	n
map	<	pre	463.88	86.75	n.s.	< 0.2	n.s.	128
		post	473.88	109.75				
	>	pre	532.04	100.36	0.003	0.21	0.009	201
		post	554.52	111.35				
GIS	<	pre	473.80	96.47	n.s.	< 0.2	n.s.	271
		post	464.24	110.31				
	>	pre	515.92	98.07	n.s.	< 0.2	n.s.	332
		post	520.89	112.51				

Tab. 75: Pre- and posttest mean values by group and pre-spatial thinking score (bin.)
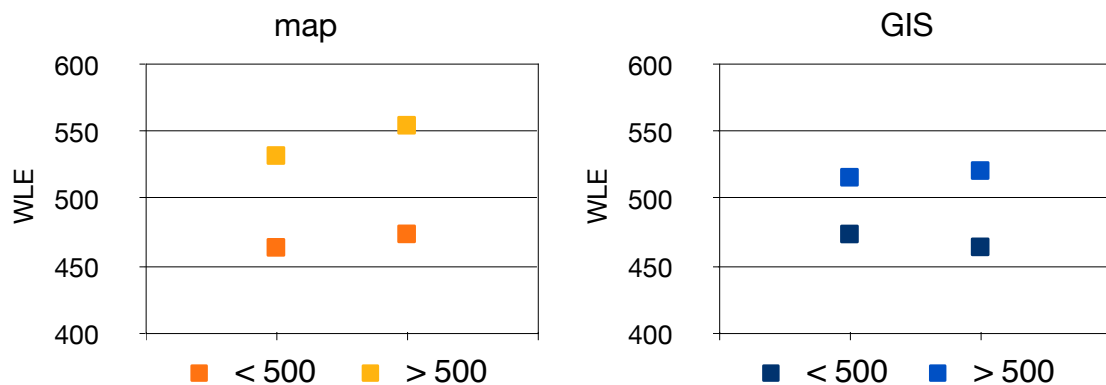
Fig. 64: Pre- and posttest mean values by group and pre-spatial thinking score (bin.)

diff WLE		M	SD	p	|d|	p (MW)	n
map	< 500	10.00	98.53	n.s.	< 0.2	n.s.	128
	> 500	22.48	105.23				201
GIS	< 500	-9.56	103.13	0.08	< 0.2	n.s.	271
	> 500	4.98	100.12				332

Tab. 76: Difference score mean values by group and pre-spatial thinking score (bin.)

diff WLE		M	SD	p	|d|	p (MW)	n
< 500	map	10.00	98.53	0.07	< 0.2	0.09	128
	GIS	-9.56	103.13				271
> 500	map	22.48	105.23	0.06	< 0.2	0.045	201
	GIS	4.98	100.12				332

Tab. 77: Difference score mean values by pre-spatial thinking score (bin.) and group
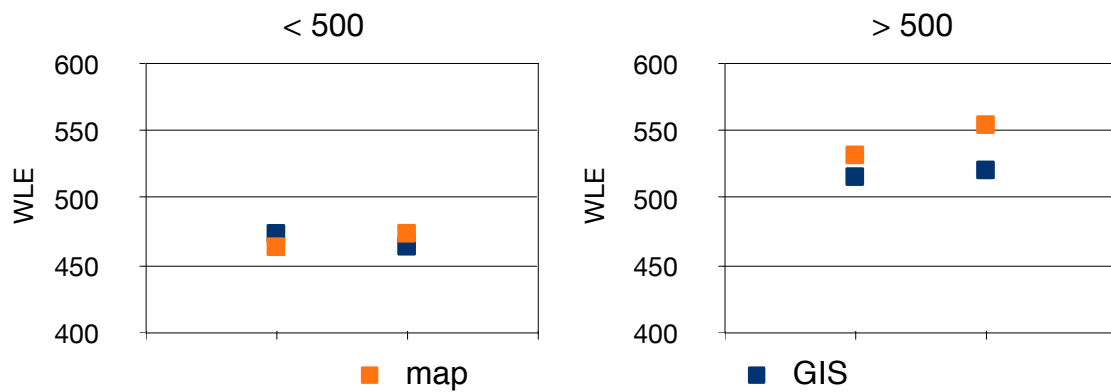
Fig. 65: Pre- and posttest mean values by pre-spatial thinking score (bin.) and group
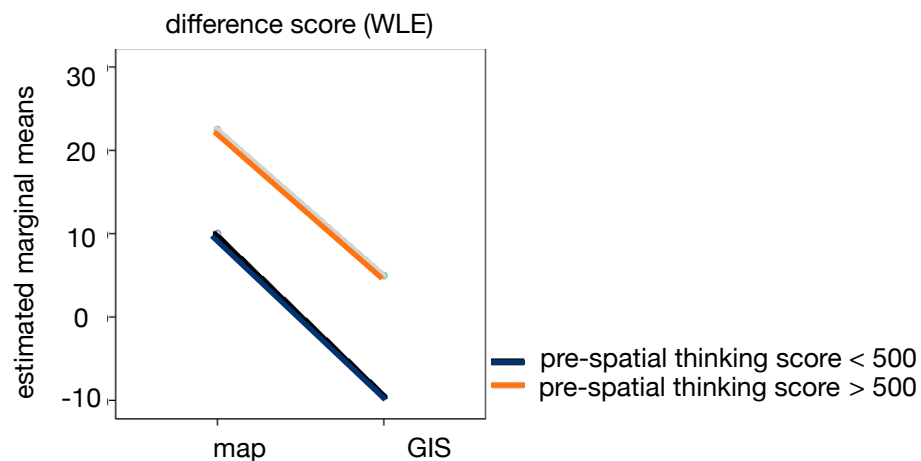
Fig. 66: Two way ANOVA graphical results: interaction group*pre-spatial thinking score (bin.)

WLE	F	p	adjusted R^2	partial η^2
group	6.79	0.009		0.01
spatial thinking pre-score bin.	3.61	0.06	0.01	0.00
group*spatial thinking pre-score bin.	0.02	n.s.		0.00

Tab. 78: Two way ANOVA: group, pre-spatial thinking score (bin.) and interaction

Pre-spatial thinking score as level 1 predictor in the HLM has no significant effect, neither does the dummy variable for the binary score (Tab. 79).

pre-spatial thinking achievement	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	6.61	n.s.	355.87	10136.85	0.003	0.00
pre-spatial	0.08	0.08				
deviance (parameters)	10514.46 (2)		homogeneity test	n.s.	reliability	0.42
intercept	-0.97	n.s.	335.61	10158.56	0.005	0.00
pre-spatial > 500	13.30	n.s.				
deviance (parameters)	10504.79 (2)		homogeneity test	n.s.	reliability	0.40
pre-spatial thinking achievement and GIS	**fixed effects**		**variance components**			**explained**
	coefficient	p	u0	R	p	
intercept	19.81	0.002	274.89	10134.66	0.02	0.00
pre-spatial	0.08	0.08				
GIS	-20.78	0.02				0.23
deviance (parameters)	10501.16 (2)		homogeneity test	n.s.	reliability	0.36
intercept	12.12	n.s.	261.52	10156.26	0.02	0.00
pre-spatial > 500	12.68	n.s.				
GIS	-20.04	0.02				0.22
deviance (parameters)	10491.79 (2)		homogeneity test	n.s.	reliability	0.35

Tab. 79: HLM: pre-spatial thinking achievement variables as level 1 predictors models

The results are mostly, but not all, in favor of the hypothesis (Tab. 80).

pre-spatial thinking score		significance	effect size
correlation between spatial thinking pre-score and systemic thinking difference score for the map group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
correlation between spatial thinking pre-score and systemic thinking difference score for the GIS group	(Tau)	✔	
	(Rho)	✔	
	(Pearson)	✔	✔
ANCOVA: pre-spatial-thinking score and group		✘ (WLE)/ ✔ (raw)	✔
HLM		✔	
pre-spatial thinking score (bin.)		**significance**	**effect size**
prepost-change within the sub-groups for the map group	(t)	✘	✘
	(W)	✘	
prepost-change within the sub-groups for the GIS group	(t)	✔	✔
	(W)	✔	
difference in prepost-change between the two sub-groups within the map group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the two sub-groups within the GIS group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the map and GIS group within the pre-spatial thinking score sub-groups	(t)	✔	✔
	(MW)	✘	
two way ANOVA: group, pre-spatial thinking (bin.) and interaction		✔	✔
HLM		✔	

Tab. 80: Overview of tests for hypothesis III-f
("On average, there is no effect of pre-spatial thinking score on improvement.")
(✘: contrary to hypothesis, ✔: in line with hypothesis)

7.2.9 Hypothesis III-g: Improvement and materials

III Improvement of achievement: On average ...

g) ... there is no effect of material type on improvement.

Only map old materials students improve significantly (*d*>0.2, Fig. 67, Tab. 81).
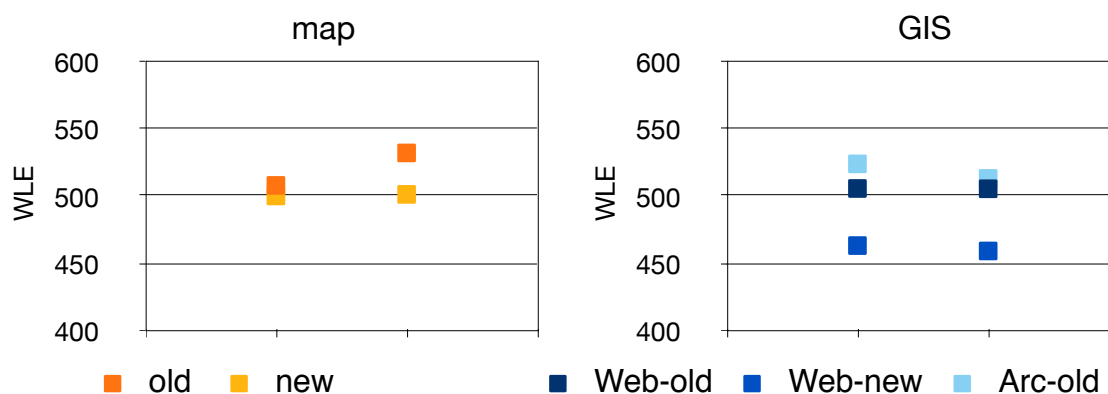
Fig. 67: Pre- and posttest mean values by group and type of materials

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
map	old	pre	0.49	0.24	< 0.001	0.25	< 0.001	507.73	104.53	0.001	0.23	0.001	236
		post	0.55	0.26				531.84	120.73				
	new	pre	0.48	0.22	n.s.	< 0.2	n.s.	499.92	90.98	n.s.	< 0.2	n.s.	93
		post	0.48	0.24				501.08	105.80				
GIS	Web-old	pre	0.49	0.22	n.s.	< 0.2	n.s.	505.46	97.15	n.s.	< 0.2	n.s.	441
		post	0.49	0.25				505.25	113.81				
	Arc-old	pre	0.53	0.24	n.s.	< 0.2	n.s.	523.71	114.79	n.s.	< 0.2	n.s.	29
		post	0.51	0.20				512.92	83.59				
	Web-new	pre	0.40	0.21	n.s.	< 0.2	n.s.	463.08	96.69	n.s.	< 0.2	n.s.	133
		post	0.39	0.25				459.08	117.81				

Tab. 81: Pre- and posttest mean values by group and type of materials

Comparing the old and new materials shows no significant differences in both map and GIS group, with the effect size being above 0.2 only within the map group in favor of the old materials (Tab. 82, Levene's test map old vs. new WLE *p*=0.09). The lack of difference in the GIS group is notable, as there thus seems to be no advantage to more 'step-by-step' materials. This remains true even when including pre-score or pre-spatial thinking score as a covariate in an AN-COVA, calculated for the WebGIS students only (Levene's test n.s., Tab. 83).

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
map	old	0.06	0.23	0.07	0.22	0.07	24.11	105.15	0.07	0.23	0.07	236
	new	0.01	0.22				1.16	94.77				93
GIS	Web old	0.00	0.23	n.s.	< 0.2	n.s.	-0.21	103.66	n.s.	< 0.2	n.s.	441
	Web new	-0.00	0.21				-4.00	96.02				133
GIS	Arc old	-0.02	0.22	n.s.	< 0.2	n.s.	-10.78	98.53	n.s.	< 0.2	n.s.	29
	Web old	0.00	0.23				-0.21	103.66				441

Tab. 82: Difference score mean values by group and type of materials

	raw				WLE			
	F	p	adjusted R^2	partial η^2	F	p	adjusted R^2	partial η^2
pre-score	90.51	< 0.001	0.13	0.14	72.57	< 0.001	0.11	0.11
Web old/new	3.84	0.05		0.01	3.76	0.05		0.01
pre-spatial-thinking-score	1.06	n.s.	-0.00	0.00	3.21	0.07	0.00	0.01
Web old/new	0.03	n.s.		0.00	0.04	n.s.		0.00

Tab. 83: ANCOVAs: pre-score/ pre-spatial thinking score and WebGIS old vs. new materials

Excluding the ArcGIS students and splitting the sample by material type shows a significant difference in favor of map students in the 'old' group ($d>0.2$, Fig. 68, Tab. 84). Comparing 'old' map and 'new' WebGIS students (Levene's test 0.08) shows a significant difference ($d>0.2$) in favor of the map group.
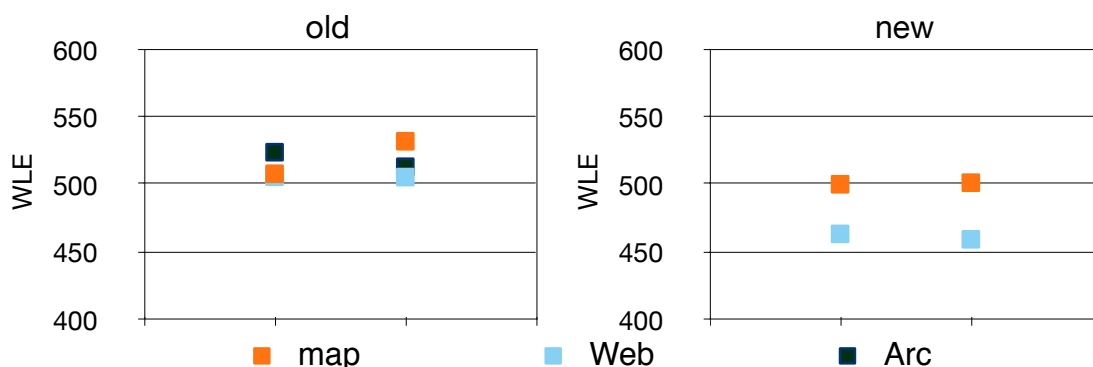
Fig. 68: Pre- and posttest mean values by type of materials and group

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
old	map	0.06	0.23	0.004	0.24	0.002	24.11	105.15	0.004	0.23	0.004	236
	WebGIS	0.00	0.23				-0.21	103.66				441
new	map	0.01	0.22	n.s.	< 0.2	n.s.	1.16	94.77	n.s.	< 0.2	n.s.	93
	WebGIS	-0.00	0.21				-4.00	96.02				133
map old		0.06	0.23	0.013	0.270	0.008	24.11	105.15	0.011	0.28	0.007	236
WebGIS new		-0.00	0.21				-4.00	96.02				133

Tab. 84: Difference score mean values by type of materials and group

Using dummy variables for the different types of materials (old map as base, thus including new map, old WebGIS and new WebGIS), there are significant negative effects for all three other treatment options compared to the map old intercept. The explained variance rises slightly to 24.2% (Tab. 85). However, the intercept chi-square is still significant, indicating that there is still class difference score variation that "[...] remains to be explained" (UCLA Statistical Consulting Group, n.d.).

material type	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	26.77	< 0.001				
map new	-24.66	0.02	268.15	10170.95	0.02	0.24
WebGIS old	-26.78	0.008				
WebGIS new	-30.91	0.01				
deviance (parameters)	10485.70 (2)		homogeneity test	n.s.	reliability	0.35

Tab. 85: HLM: material types as level 2 predictors model

The results are mixed (Tab. 86), consequently, the hypothesis should be rejected. It is interesting to note that there is no significant difference in the new materials and that there are no significant differences within the GIS group.

material type		significance	effect size
prepost-change within the sub-groups for the map group	(t)	✗	✗
	(W)	✗	
prepost-change within the sub-groups for the GIS group	(t)	✔	✔
	(W)	✔	
difference in prepost-change between the two sub-groups within the map group	(t)	✔	✗
	(MW)	✔	
difference in prepost-change between the sub-groups within the GIS group	(t)	✔	✔
	(MW)	✔	
difference in prepost-change between the map and WebGIS group within each material type sub-group	(t)	✗	✗
	(MW)	✗	
ANCOVAs with pre-score/ pre-spatial-thinking-score and WebGIS old vs. new materials		✔	✔
HLM		✗	

Tab. 86: Overview of tests for hypothesis III-g
        ("On average, there is no effect of material type on improvement.")
        (✗: contrary to hypothesis, ✔: in line with hypothesis)

7.2.10 Hypothesis III-h: Improvement and pre-experience

III Improvement of achievement: On average, ...

> *h) ... pre-experience with an educational GIS program has a positive effect on improvement.*

There are only 23 students that can be included as having specified to have already worked with an educational GIS. Excluded are those answering "no" but specifying "earth" (or similar), those answering "yes" but nothing or those answering "yes" but not a specific educational GIS (e.g. Google Earth or a satnav). Only students that have matchable materials are included.

Neither the students with nor those without pre-experience with GIS change significantly pre- to posttest (Fig. 69, Tab. 87). However, the effect sizes for the students with pre-experience are above 0.2.
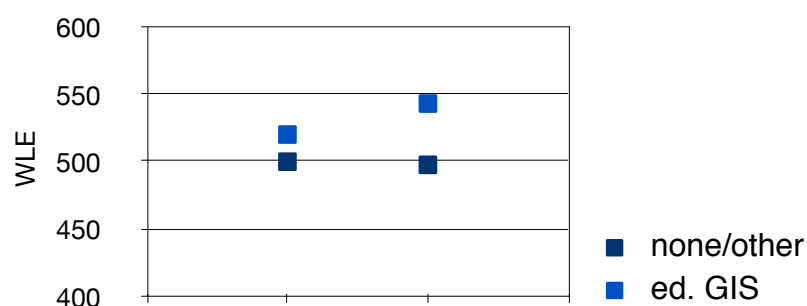
Fig. 69: Pre- and posttest mean values of the GIS students by pre-experience

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
GIS	none/ other	pre	0.48	0.22	n.s.	< 0.2	n.s.	499.75	98.26	n.s.	< 0.2	n.s.	520
		post	0.48	0.25				497.40	114.84				
	ed. GIS	pre	0.53	0.17	n.s.	0.27	n.s.	519.90	71.82	n.s.	0.27	n.s.	23
		post	0.58	0.20				542.91	81.23				

Tab. 87: Pre- and posttest mean values by pre-experience (GIS group)

The difference in achievement change is not significant, however, the effect sizes are above 0.2, in favor of the students with experience (Tab. 88).

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
GIS	none/other	-0.00	0.23	n.s.	0.26	n.s.	-2.35	103.72	n.s.	0.27	n.s.	520
	ed. GIS	0.06	0.21				23.01	85.98				23

Tab. 88: Difference score mean values by pre-experience (GIS group)

The sample is heavily imbalanced in terms of sample size. Those students included in the 'educational GIS pre-experience'-sub-group are only in the high stream and only from Baden-Württemberg. This is interesting in itself, pointing to an even lower exposure of mid stream students than high stream students. Two of the students are in the ArcGIS group, the others used WebGIS. However, two students can be hardly used for a comparison. Consequently, only WebGIS students could be compared, leaving 183 students in the 'none/other' and 21 in the 'educational GIS pre-experience' category, which is still imbalanced, but less than originally. There is only one student with migration background in the latter category. Consequently, excluding students with migration background leads to 149 and 20 students respectively. One of the remaining ed. GIS students is twelve, all the others are 13. Looking only at the 13-year-olds gives a sample of 84 and 19 respectively. Pre-post changes are not significant, but students with pre-experience have a small effect size (Fig. 70, Tab. 89, 90).
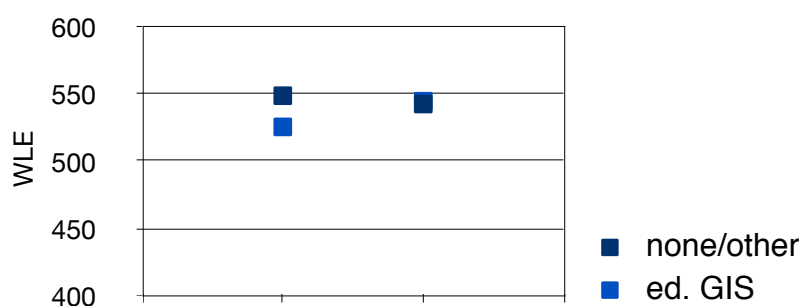
Fig. 70: Pre- and posttest mean values of the GIS students (selection) by pre-experience

			raw		p	|dz|	p (W)	WLE		p	|dz|	p (W)	n
			M	SD				M	SD				
GIS	none/other	pre	0.59	0.22	n.s.	< 0.2	n.s.	548.49	97.98	n.s.	< 0.2	n.s.	84
		post	0.57	0.25				542.32	116.29				
	ed. GIS	pre	0.54	0.18	n.s.	0.20	n.s.	525.34	75.33	n.s.	0.21	n.s.	19
		post	0.59	0.21				544.40	87.77				

Tab. 89: Pre- and posttest mean values by pre-experience (GIS group selection)

		diff raw		p	|d|	p (MW)	diff WLE		p	|d|	p (MW)	n
		M	SD				M	SD				
GIS	none/other	-0.01	0.25	n.s.	0.24	n.s.	-6.17	112.80	n.s.	0.24	n.s.	84
	ed. GIS	0.05	0.22				19.05	92.10				19

Tab. 90: Difference score mean values by pre-experience (GIS group selection)

Calculating a HLM for this hypothesis does not make sense due to the insufficient sample size when looking only at the GIS students.

material type		significance	effect size
prepost-change within the sub-groups for the GIS group	(t)	✗	✔
	(W)	✗	
difference in prepost-change between the sub-groups within the GIS group	(t)	✗	✔
	(MW)	✗	
prepost-change within the sub-groups for the reduced GIS group	(t)	✗	✔
	(W)	✗	
difference in prepost-change between the sub-groups within the reduced GIS group	(t)	✗	✔
	(MW)	✗	

Tab. 91: Overview of tests for hypothesis III-h
("On average, pre-experience with an educational GIS program has a positive effect on improvement.") (✗: contrary to hypothesis, ✔: in line with hypothesis)

The achieved power is very low because the sample size for students already having pre-experience with educational GIS is very small. While the results are not significant, effect sizes (Tab. 91) hint at a potential positive influence of pre-experience and thus possible support for the hypothesis with larger samples.

7.2.11 Overall HLMs

stream and GIS	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	6.76	n.s.				
high stream	19.09	0.01	200.06	10178.29	0.054	0.43
GIS	-20.03	0.01				
deviance (parameters)	10489.87 (2)		homogeneity test	n.s.	reliability	0.29
stream and material type	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	12.95	n.s.				
high stream	17.18	0.06				
map new	-17.42	n.s.	219.59	10173.89	0.047	0.38
WebGIS old	-25.60	0.005				
WebGIS new	-23.02	0.10				
deviance (parameters)	10474.29 (2)		homogeneity test	n.s.	reliability	0.310

Tab. 92: HLM: level 2 predictors models

Combining the analysis for level 2 shows a significant negative effect for GIS and a significant positive effect for high stream. 43.5% of class variation can be explained by GIS and high stream together, with the chi-square value becoming non-significant. Only the GIS old effect (negative) is still significant. The percentage of class variation that can be explained decreases to 38.0%, with the chi-square still being significant (Tab. 92).

Combining all level 1 variables leads to a significant homogeneity test. All together, 19.8% of the variance is explained, with a significant positive effect for pre-spatial-thinking-score and negative effects for pre-score and male (Tab. 93).

all level 1	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	16.97	0.006				
pre	-0.48	< 0.001				
pre-spatial	0.17	< 0.001				
male	-18.13	0.01	437.05	8161.23	< 0.001	0.20
age	-5.26	n.s.				
migration	10.72	n.s.				
total language	-8.53	n.s.				
family language other	-22.70	0.08				
deviance (parameters)	10303.89 (2)		homogeneity test	0.006	reliability	0.52

Tab. 93: HLM: level 1 predictors model

Deleting non-significant effects again has the problem of a significant homogeneity test. All together, 19.2% of the variance is explained, with a significant positive effect for pre-spatial-score as well as negative effects for pre-score and male (Tab. 94). Consequently, the other level 1 variables added less than 1% to the variance explained.

significant level 1	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	15.41	0.008				
pre	-0.48	< 0.001	405.34	8224.82	< 0.001	0.19
pre-spatial	0.17	< 0.001				
male	-17.76	0.02				
deviance (parameters)	10336.79 (2)		homogeneity test	0.004	reliability	0.50

Tab. 94: HLM: level 1 predictors model (selection)

Combining all level 1 variables with binary scores for pre- and pre-spatial thinking achievement, all together 10.5%, of the variance gets explained with a significant positive effect for pre-spatial-achievement as well as negative effects for pre-score and male (Tab. 95).

level 1 binary	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	29.10	0.001	538.71	9105.86	< 0.001	0.11
pre > 500	-61.93	< 0.001				
pre-spatial > 500	25.66	0.002				
male	-16.97	0.02				
age	-7.85	n.s.				
migration	14.16	n.s.				
total language	-6.75	n.s.				
family language other	-24.79	0.08				
deviance (parameters)	10379.44 (2)		homogeneity test	n.s.	reliability	0.55

Tab. 95: HLM: level 1 predictors model (bin. scores)

Deleting the non-significant effects, all together 9.9% of variance is explained, with a significant positive effect for pre-spatial score and negative effects for pre-score and male (Tab. 96). Again, the remaining variables had added less than 1% of the variance explained.

level 1 binary signifi-cant only	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	27.33	0.001	508.91	9171.76	< 0.001	0.10
pre > 500	-60.72	< 0.001				
pre-spatial > 500	25.79	0.002				
male	-16.92	0.02				
deviance (parameters)	10412.65 (2)		homogeneity test	n.s.	reliability	0.53

Tab. 96: HLM: level 1 predictors model (bin. scores, selection)

level 1 and level 2	fixed effects		variance components			explained
	coefficient	p	u0	R	p	
intercept	14.88	0.05	268.12	8224.67	0.006	0.19
GIS	-18.95	0.02				
high stream	18.57	0.02				
pre	-0.48	< 0.001				0.34
pre-spatial	0.17	< 0.001				
male	-17.20	0.02				
deviance (parameters)	10314.96 (2)		homogeneity test	0.004	reliability	0.40

Tab. 97: HLM: level 1 and level 2 predictors model (selection) with GIS

Combining level 1 and level 2 variables (non-binary) leads to a significant homogeneity test. All together, 19.2% of level 1 and 33.9% of level 2 variance are explained, with still significant variance left (Tab. 97). There are significant

negative effects for GIS, pre-score and male as well as positive effects for pre-spatial-score and high stream.

Combining level 1 and level 2 variables (non-binary) using type of material leads to a bit larger differences and a significant homogeneity test. There are 19.2% of level 1 and 29.6% of level 2 explained, with still significant variance left (Tab. 98). There are significant negative effects for GIS old, pre-score and male and positive effects for high stream and pre-spatial-score.

	fixed effects		variance components			variance explained
	coefficient	p	u0	R	p	
intercept	21.02	0.04				
map new	-17.33	n.s.				
WebGIS old	-24.58	0.007				0.19
WebGIS new	-21.87	0.09				
high stream	16.69	0.06	285.43	8221.33	0.005	
pre	-0.48	< 0.001				
pre-spatial	0.17	< 0.001				0.30
male	-17.09	0.02				
deviance (parameters)	10299.41 (2)		homogeneity test	0.004	reliability	0.42

Tab. 98: HLM: level 1 and level 2 predictors model (selection) with material type

Repeating the analysis with binary scores and GIS leads 9.9% of level 1 and 40.2% of level 2 variance is explained, with still significant variance left. There are significant negative effects for GIS, high pre-score and male, and positive effects for high stream and high spatial thinking (Tab. 99).

	fixed effects		variance components			variance explained
	coefficient	p	u0	R	p	
intercept	23.01	0.009				
GIS	-17.94	0.04				0.10
high stream	28.19	0.003				
pre > 500	-62.66	< 0.001	304.10	9165.01	0.006	
pre-spatial > 500	21.55	0.01				0.40
male	-16.28	0.03				
deviance (parameters)	10387.72 (2)		homogeneity test	n.s.	reliability	0.40

Tab. 99: HLM: level 1 and level 2 predictors model (bin. scores, selection) with GIS

Repeating the analysis with binary score and material type 9.9% of level 1 and 33.0% of level 2 variance is explained, with still significant variance left. There are significant negative effects for high pre-score, GIS old and male, and sig-

nificant positive effects for high stream and high spatial thinking (Tab. 100). Interestingly, the negative effect for GIS new is not significant.

	fixed effects		variance components			variance explained
	coefficient	p	u0	R	p	
intercept	25.84	0.02	340.72	9164.51	0.004	0.10
map new	-8.05	n.s.				
WebGIS old	-20.47	0.04				
WebGIS new	-19.62	n.s.				
high stream	27.20	0.01				
pre > 500	-62.58	< 0.001				
pre-spatial > 500	21.58	0.01				0.33
male	-16.18	0.03				
deviance (parameters)	10373.04 (2)		homogeneity test	n.s.	reliability	0.43

Tab. 100: HLM: level 1 and level 2 predictors model (bin. scores, selection) with material type

Looking at the combined analyses shows again no significantly positive effect for GIS. Interestingly, while GIS overall has a negative effect, there are difference in terms of the materials once the other variables are included, namely, only the negative effect of the old WebGIS materials is significant, while that of the new map materials and the new WebGIS materials is not. Consequently, here it would appear that more step-by-step materials are somewhat of an advantage. This stands in contrast to the earlier analysis (see hypothesis III-g).

The HLMs show no effect of migration and language background as well as age. In contrast, there is a consistent negative effect for male, as well as a consistent positive effect of good pre-spatial-thinking-scores and attending a high stream school. The consistent negative effect of high pre-systemic-thinking-scores might be due to test characteristics, i.e. a potential ceiling effect, and should thus be examined in further studies.

Overall, in both complex HLMs, there is still significant variance left to be explained, i.e., not all relevant variables seem to have been included in this study. This is especially noticeable in the case of level 1, i.e. student level, variance.

7.2.12 Conclusions from Study 5

The main hypothesis could not be supported. On average, GIS students do not improve pre- to posttest in systemic thinking, albeit there are large differences

between classes. Thus, overall, GIS has no positive, and partly a significant negative impact compared to maps.

Interesting is the difference between the WebGIS old and new materials, albeit the imbalance of the sample (see ch. 4) should be kept in mind. For instance, whereas t-tests show no significant difference between old and new WebGIS materials, in the complex HLM only the old variety has a significantly negative effect. This indicates that either a more step-by-step could be potentially better for the students when dealing with GIS or that this could also be simply a threshold problem. Consequently, this should be explored further.

Moreover, effect sizes hint at students with pre-experience, which are very few in number, achieving better results than students without. This might also be an explanation for the bad GIS results, as students might have had too much cognitive capacity captured by learning to handle the software and thus less to focus on the system connections.

Other variables showed mixed results. Moreover, the HLMs show that not all relevant variables have been included in the study.

8 Discussing the implications

The studies conducted for this dissertation contribute to two areas of geography education in Germany that have still many areas to be explored: systemic thinking and GIS use. They also provide a contribution to spatial thinking research.

Of the five studies conducted, study 5 achieved a sample of 932 students for the analysis, which is comparatively large both for a pre-/posttest geo-science education study in Germany (e.g. Ditter, 2013, n=322; Fiene, 2014, n=295) and a GIS in education study (e.g. Aarons, 2003, n=84; Baker, 2002, n=170; Crews, 2008, n=52; Falk & Nöthen, 2005, n=16; Lee, 2005, n=80). It is also an adequate size compared to other systemic thinking studies (e.g. Ben-Zvi Assaraf & Orion, 2005a, n=70; Ben-Zvi Assaraf & Orion, 2005b, n=1000; Ben-Zvi Assaraf & Orion, 2010b, n=40; Kali, et al., 2003, n=40; Orion & Basis, 2008, n=31; Ossimitz, 1996, n=126; Siegmund, et al., 2012, n1=110, n2=324; Sommer, 2005, n=363).

There have been many positive claims surrounding GIS (see ch. 2) and results in favor of GIS reported by other studies (e.g. Aladag & Aladag, 2008; Baker & White, 2003; Kerski, 2000; Lee, 2005; Patterson, et al., 2003; Wanner & Kerski, 1999). These studies have been, however, conducted with different instruments, different treatments, a different target outcome, a different sample (also in size) and mostly in different countries. In both study 3 and study 5 conducted for this dissertation, GIS students do not improve significantly. While in study 3, with twelfth grade students, the pre-post-change is positive ($d>0.2$), in study 5, with seventh grade students, the values stay nearly the same ($d<0.2$). In both studies, there is no advantage of GIS compared to maps. In study 5, the difference between map and GIS students is actually significant in favor of map students ($d<0.2$). There are other studies such as Meyer et al. (1999) who found that the control group scored better than the GIS group on a location analysis task. Moreover, Klein (2007) found that among the media studied, GIS was the one with the highest percentage of students reporting that they understood nothing when working with it. Falk and Nöthen (2005, p. 51, translated) also state that their GIS project, in a study in Germany, "[...] could only moderately contribute to fostering systemic thinking". Non-GIS related studies of systemic thinking in

Germany not showing improvement in all or some of the groups include Rieß and Mischo (2008b).

8.1 Implications based on the student variables included

The reasons for non-significant improvements can be multifaceted (see also ch. 2). For instance, Orion and Basis (2008), based on an Israeli study, argue that "[...] the initial cognitive ability of the students, the curriculum that was built as a result of the outdoor learning experience, knowledge integration activities, the teachers' mediation and guidance, the teachers' emotional support to the students, the students' level of involvement, and the students' perception of the learning process" are all important variables and that "[s]tudents develop systems thinking skills only when all these factors synergize". Hattie (2003) states that, in general, student factors "[...] account for about 50% of the variance of achievement" (p. 1), home, school and peer factors each "[...] account for about 5-10% [...]" (p. 2) and teachers "[...] account for about 30% [...]" (p. 2). A later published meta-analysis includes over 100 factors (Hattie, 2009).

In study 5, several variables are included, many of which show mixed results. However, the HLM indicates that there is still variance left to be explained, i.e., not all variables that seem to have effects are included in the study. Some variables that were included in the earlier studies, such as interest or pre-experience variables e.g. regarding computers (see Tab.1, ch. 3), had to be excluded in study 5 due to test time considerations, but should be tested in future studies. Moreover, additional examples for candidates for possibly influencing variables are discussed below. However, it has to be kept in mind that individual studies have limits on the number of variables that can be included.

8.1.1 Age

Within one grade, in study 5, ANOVAs for age are only significant within the GIS group, with the older the students the worse the pre-posttest change (effect size below RMPE). Further analyses show mixed results. The overall HLM does not show a significant effect of age. Because only one grade level was analyzed in study 3 and study 5 each, no comparisons across a broad age range can be made.

In general, grade seven was selected due to curricular and organizational reasons (see ch. 3). In terms of systemic thinking achievement, in a study with seventh to ninth grade students, Ben-Zvi Assaraf and Orion (2005b, p. 368) found "[...] no significant difference [...] between grades regarding any of the research tools [...]." Thus, grade seven seemed suitable, despite grade seven students also being in the midst of puberty.

However, interest studies showed that in general, younger and older students seem to have positive characteristics with regard to interest, while for instance with regard to human geography, grade seven students scored lowest (see overview in Hemmer & Hemmer, 1997; Hemmer & Hemmer, 2002a). Ditter (2013) found significant effects for age, e.g in terms of motivation/self-efficacy/interest cluster composition in a study including classes from grades seven to twelve, however, these were not positive throughout for elder students. Besides for instance differences in the materials and test instruments as well as students' experience, affective variables thus could be one of the possible reasons for the positive (d>0.2), albeit not significant change of GIS students in study 3 (with older students) compared to the lack thereof in study 5. Although the exclusion of the affective variable scales in study 5 was necessary due to test time considerations, they should be included in future studies in order to test for a possible influence of affective variables.

8.1.2 Sex

In both study 3 (t-test for the GIS group, d>0.2) and study 5 (overall in the HLM; and in the t-test for the GIS group, d>0.2) there is a significant effect of sex in favor of females. This is surprising considering that males are often shown to be on average more tech savvy than females, or at least self-describe more confidently (e.g. Klein, 2007; Obermaier, 2005; OECD, 2011) and have "[...] more positive attitudes towards computers than girls [...]" (OECD, 2011, p. 168). They also use computers more for leisure activities (ibid., p. 159). Males also often have a better map competency (e.g. Hüttermann, et al., 2010) and a better satellite image reading literacy (Kollar, 2012). Moreover, males have often, albeit far from always, better achievement in earth science and systemic thinking (see ch. 2). Moreover, the results of studies 3 and 5 are in contrast to many of the GIS studies showing no significant effects of sex (see ch. 2.3.2), but in line with those of

Clark et al. (2007), who also found a higher achievement of females in a GIS group. In a study of the effects of working with digital remote sensing data on students' motivation and self concept, Ditter (2013) also found positive effects for females, albeit from a lower starting value than males. In general, male students are more frequently trying to avoid effort (e.g. summary in Rollett, 2011). Consequently, further studies could include control variables such as effort and computer skills to get a more detailed look at possible reasons for the contrasting results. Furthermore, some research has shown an effect of task format on sex differences in achievement (see e.g. summary in Bednarz, et al., 2013).

8.1.3 Migration and language background

The results in terms of migration and language background are mixed. For instance, there are no significant effects of migration background, total number of languages or family language background in the HLM. Yet, in the sub-group 'students with migration background' there is no significant difference between map and GIS students (d<0.2), while in the no migration background students there is one (in favor of map students, d>0.2; similar when taking family language background). This could point to GIS having a somewhat positive effect for students without migration background but not for students with migration background, making the scores more similar. However, in the map group, there is no significant difference between students with and without migration background (d<0.2).

There are several possible explanations for the studies showing no pronounced negative effect for migration background. One is that while in the earlier PISA-tests, in Germany "[...] adolescents from immigrant families performed substantially worse than adolescents without migration background in all areas", PISA 2012 showed that "[t]he competencies of the adolescents with migration background have improved considerably and consistently [...]" (Göroğlu, 2013, translated). Another possible explanation is that the treatment was too short to make potential differences more visible. A third possible explanation is that the results are due to the sample composition. Statistics about the percentage of students with migration background attending the different streams show considerable differences by country (e.g. Beauftragte der Bundesregierung für Migration Flüchtlinge und Integration, 2005; Burgmaier & Traub, 2007; Geißler &

Weber-Menges, 2008; Glorius, 2012; Stutzer, 2005). For instance, compared to students with German citizenship, students with Russian citizenship go to high stream schools about as frequently, while students with Turkish or Italian citizenship do so substantially less.

A further exploration of the variables migration and language background would thus require even larger sample sizes in order to be able to take a more detailed look, e.g. at different sub-groups. Another avenue would be to include the social economic background as a controlling variable (see e.g. Geißler & Weber-Menges, 2008; OECD, 2006b, 2007c).

Additionally, student answers sometimes indicated problems with language. Since the method of having students dictate answers to the researchers used e.g. by Ben-Zvi Assaraf and Orion (2010b) is not feasible with large sample sizes, future studies could include e.g. a reading literacy test or the last language grades, both in German and foreign languages, as possible influencing variables, instead of the general language background question.

8.1.4 Pre-achievement and pre-spatial thinking

Different pre-achievement measures were examined, namely pre-score, pre-spatial-thinking-score, and, in study 5, also stream. Pre-achievement influences have been well documented in a variety of areas (see ch. 2), albeit they do not always occur (see e.g. Hildebrandt, 2006). The results of studies 3 and 5 support the importance of including pre-achievement variables when studying the effects of a treatment on systemic thinking.

Firstly, in terms of pre-score, both studies mostly show significant effects. For instance, within the GIS group, t-tests are in favor of low scoring students ($d>0.2$) in both studies, albeit in study 3, the difference is not significant. Similarly, in the study 5 HLMs, there is a significant negative effect for pre-score. This could point to possible ceiling effects needing to be further explored or to the unit being more suitable for low pre-achievement students. The results are partly in line with those reported by Kerski (2000) who found significant effects with students "[...] with lower pretest scores improv[ing] the most", looking at the influence of spatial analysis pre-scores on spatial analysis difference scores (p. 221). However, in contrast, he found no significant effects of standardized pre-scores on

standardized difference scores. Maierhofer (2001), in a biology study dealing with systemic thinking, also found that while students' scoring low on the system variable pretest improved significantly, high scoring students did not.

Secondly, in terms of pre-spatial-thinking-score, study 5 shows somewhat mixed results, but mostly no significant effects in the individual analyses. However, in the overall HLM with other variables included, there is a significant positive effect for high pre-spatial thinking achievement, which could hint at those basic spatial thinking skills being necessary – or at least helpful – to work with the unit. The relationship between spatial and systemic thinking needs to be further explored, both in terms of competency models (see e.g. the HEIGIS project described in ch. 2) and the possible influences of spatial thinking skills on achievement improvement. Because basic spatial thinking skills have been also included in the posttest, later analyses of the data set could also explore their improvement through GIS vs. map use.

Thirdly, the results for stream in study 5 are mixed. In the GIS group, there is only a significant difference between mid and high stream students in the WLE analysis ($d<0.2$, in favor of high stream students). While there is no significant difference between map and GIS students in the mid stream, there is one in the high stream (in favor of map students, $d<0.2$). Similar to existing research (see ch. 2), in general, this hints at possible school type differences. In many of the HLMs there is a significant positive effect for high stream overall.

8.1.5 Pre-experience with GIS

Despite the comparatively large sample in study 5, there were only very few students that had pre-experience with GIS. Thus, its influence can only be analyzed in a very exploratory manner. Although the differences in improvement between students with and without pre-experience are not significant, the effect sizes are above 0.2. Student feedback hints at the complexity and unfamiliarity of the medium as being a problem. Similarly, Kollar (2012), for instance, showed a partly positive effect of overall pre-experience with satellite images on satellite image reading literacy. This could be explained for instance with the help of cognitive load theory (see e.g. overview in de Jong, 2010; and discussion in Sell, et al., 2006).

Thus, in further studies, both pre-experience as a variable and further control variables (e.g. working memory capacity, see de Jong, 2010) should be further explored (see e.g. also Hildebrandt, 2006). As GIS use becomes more frequent, it should become easier to find classes that have already worked with GIS. Another possibility would be to include a pre-training in GIS before starting with the actual content unit. Moreover, the GIS competency that the students had before or acquired in the course of the unit has not been measured. The inclusion of this variable would aid testing the hypothesis that cognitive load could have played a role in the lack of improvement of GIS students.

8.2 Implications based on sample size and analyses used

The sample size is fairly large by GIS treatment study standards (see above). Yet, it is still small when looking at sub-groups, limiting the conclusions that can be drawn. In general, further studies with larger samples would be helpful (see also e.g. Bednarz, et al., 2013). Moreover, the sample is not perfectly balanced. This is a general problem of studies that need to rely on the voluntary participation of students and teachers. Only analyzing a randomly drawn sub-sample of the available data based on certain criteria in order to be able to use a more balanced sample (see e.g. Obermaier, 1997) seemed not sensible at the present sample size level.

The analyses were conducted with the help of both the frequently used raw scores and the in the past much less, but now increasingly used IRT scores, which have various advantages (see ch. 3). Moreover, the analyses used included both those used frequently in geography education studies such as t-tests, but also the hitherto seemingly not often used HLMs. HLMs take the nested structure of class/student samples into account and are thus a potentially important tool for analyzing data in geography education (see ch. 3). However, the sample size of study 5 (44 classes) is very close to the lowest sample size requirements of HLMs. Together with the issues of the measurement instrument used and the sample composition, this means the results need to be treated with some caution. Moreover, it also means that separate analyses of the map and GIS group as well as the study of interactions of effects within the whole sample were not feasible.

The results highlight the importance of the analyses chosen, especially in terms of controlling for other factors. For instance, in study 5, both in a t-test ($d<0.2$) and HLMs with and without other variables included there is an overall significant negative effect for GIS. An analysis by material type shows significant negative effects for WebGIS old but not new in HLMs with other variables included. However, the HLM with only material type as variable produces negative effects for WebGIS old, WebGIS new and map new materials and t-tests show no significant difference between WebGIS old and new materials ($d<0.2$). Thus, whether more 'step-by-step' ('new') WebGIS materials are advantageous compared to providing students only with a summary sheet explaining the basic functionality ('old') seems to be partly a matter of the analysis method used to examine the data. Moreover, the sample is fairly imbalanced (e.g. significantly lower pre-score for WebGIS new compared to WebGIS old students at $p<0.001$ for both WLE and raw). Consequently, to further examine these issues larger (and more balanced) samples should be used.

8.3 Implications based on the treatment used

The treatment duration (four lessons) was fairly short. However, earlier GIS in school studies used both longer (e.g. Baker & White, 2003, nine days) and shorter (e.g. Demirci, 2008, 40 minutes) units. The duration of the present study was similar to that of the remote-sensing lessons used by Ditter (2013, four lessons).

Student feedback to the unit was mixed in both groups. Several students, especially of the GIS group, stated that they found GIS hard at the beginning and had to get used to it (see ch. 6).

8.3.1 Material type

Study 5 used two different types of learning materials for the unit, whereby the 'new' material type was only represented in the WebGIS, not the ArcGIS group. The two types of materials used parallel questions, but different ways of technical help. Results with regard to material type are mixed based on the analysis method used (see ch. 8.2). On the one hand, step-by-step materials can help students, especially those that are not yet familiar with GIS. This kind of material

is often used (e.g. Keranen & Kolvoord, 2008; Palmer, et al., 2008a, b; Palmer, et al., 2008c, d; Püschel, 2011; Püschel, et al., 2007). On the other hand, in between step-by-step explanations also 'break the flow' from dealing with the content, especially for those students that would be able to do without (compare e.g. the results regarding the "[…]"expertise-reversal effect"[…]" summarized in de Jong, 2010, p. 111). Working with the new WebGIS materials did not have a significantly negative effect compared to the old WebGIS materials and in the HLMs partly did not have a significant effect, while the old WebGIS materials did. Consequently, for studies wanting to work only with one type of materials, it seems to be overall more advisable to use step-by-step materials for samples with a high percentage of students without GIS pre-experience.

There were only very few students working with ArcGIS. Consequently, no substantial comparisons can be made, especially due to some methodological problems with part of the sample (see ch. 4). A t-test between the old material ArcGIS and the corresponding WebGIS students shows no significant difference ($d<0.2$). The ArcGIS students were excluded from the HLMs.

Overall, there were very few students with pre-experience with GIS (see ch. 8.1). Thus, it remains unclear whether a simple Web-based GIS already could have a positive impact, e.g. for students with more pre-experience, after a GIS training or in a different setting, or whether e.g. a professional DesktopGIS is needed. Moreover, it is not yet fully understood which features of a GIS are especially beneficial (or hindering) for improving systemic thinking skills.

8.3.2 Methodology

Study 5 used learning materials for the unit that are alike for all classes to ensure comparability. As discussed, this might lead to the unit not being a fit for a specific class situation. Another possibility would be to have a large material pool with different degrees of difficulty and utilized methods from which the teachers and/or students could choose. However, this would not completely eliminate the situation 'fitting to the class situation' and success would at least somewhat depend on the skills/pre-experience to choose fitting materials. Moreover, in such studies, differentiation should be one of the explicitly studied

research questions, as due to the then very large matrixes it seems improbable to achieve sufficient sample sizes for individual material comparisons.

Additionally, the materials called for students working (a) largely on their own, without teacher instruction and (b) together with a partner. This was done to limit instructor effects and due to likely school computer lab size, but might both be difficult for some students. Since variables such as independent study pre-experience or attitude to partner work were not included, statements about proportions of students having problems with this and their possible influence on study outcome cannot be made.

Overall, it should be noted that even in such situations, instructor effects cannot be completely eliminated, e.g. in terms of their skills and confidence in answering student questions or their attitude towards the tests before and after the unit. For instance, Ditter (2013, p. 184) observed considerable effects of regular teacher's interest on student motivation change pre- to posttest despite teaching the treatment unit himself in all classes. For the studies conducted within this study, due to the large sample size, teaching or even being present in all classes was not feasible. Orion and Basis (2008) name "[...] the teachers' mediation and guidance, [and] the teachers' emotional support to the students [...]" as two of the factors contributing to the development of systemic thinking. Additionally, Hattie (2003, 2009), Weinert (2001) and Carneiro (2006) also highlight the importance of the teacher.

While teachers in study 5 could participate in a short introduction to GIS if they wanted, no data was collected regarding the teachers' actual GIS/map competency, pre-experience and attitude towards GIS/map use. This could be important for the effects of GIS use, however, e.g. due to the flexibility and ability with which a teacher can support students in the case of (technical) problems. Thus, these variables could be included in future studies.

There also has been no process assessment (e.g. how involved the students were in the process). This could be an important variable, as Orion and Basis (2008) also include "[...] the students' level of involvement, and the students' perception of the learning process" as important factors to develop systemic thinking (see also e.g. Ben-Zvi Assaraf & Orion, 2010b). Such assessment could be done by observation,

but this would likely put considerable limits on sample size. Consequently, self-assessment questions could be used as an approximation, e.g. by including control questions based on the FAM (survey of current motivation, see e.g. Rheinberg, et al., 2001), attitude/interest questions regarding different aspects (e.g. the topic, the medium, see ch. 4) as well as questions such as "[...] "In todays research lesson/ mathematics lesson we have worked concentratedly"" (Peer, et al., 2008, p. 25, translated) sometimes used in school practice (see also e.g. Hildebrandt, 2006). For study 5, attitude/interest questions e.g. towards the topic and medium could not be included due to test time considerations.

All in all, there are many different ways to design the learning materials, choose questions, etc. to help students explore the topic even using the same WebGIS service. While the materials in study 3 and study 5 had not yet been tested in one or several rounds at the time of their use in a pre-posttest study, the development was theory-guided and drafts had been discussed with experts, such as experienced teachers (see ch. 6). By now, the materials have been further developed by the team of the "GIS Station – Klaus Tschira Competency Centre for Digital Geomedia" at the Heidelberg University of Education. Additionally, even at the time of the studies there had been spin-off materials also dealing with tourism in Kenya (Viehrig, 2009; Viehrig & Volz, 2009). The choices made when designing the learning materials have a potentially great effect on the outcome, both directly e.g. by helping students understand certain relationships better, and indirectly by influencing factors such as motivation and self-concept which in turn have effects on achievement (for the later see e.g. overview in Ditter, 2013). Consequently, for future projects, design-based research methods (e.g. Institut für Physik und ihre Didaktik, 2013; Orion & Cohen, 2007), seem a promising avenue for learning material design before or during studying its impact.

Moreover, Ben-Zvi Assaraf and Orion (2010b) found that "[...] students could better identify relationships between components they explored during the outdoor learning activities than those which were not explored in concrete contexts" (p. 559). Thus, a combination of GIS and concrete experience would also be an interesting avenue to be studied.

8.4 Implications based on the test instruments used

Despite the importance of systemic thinking in the geography education discourse there is still a lack of comprehensive, empirically validated competency models and test instruments for a variety of topics, ability levels and age ranges (see e.g. ch. 2 and Siegmund, et al., 2012). The studies conducted for this dissertation contribute to the growing amount of literature addressing this lack and especially to test aspects of the model developed by Orion and colleagues for a different topic and sample. To provide further ideas and results for developing testing instruments for systemic thinking in geography is very necessary because test instruments are the foundation for the empirical examination of other geographical didactical questions. As Lichtenstein et al. (2008, p. 610) argue for science education:

> "Only with validated instruments will education researchers be able to move forward to determine if curricular and instructional innovations truly improve student attitudes toward science".

This also means, however, that to some extent the results of this study regarding the effects of GIS need to be treated cautiously and seen as preliminary.

8.4.1 Systemic thinking

In all studies where systemic thinking tasks were included, a Rasch scale could be achieved in the end, e.g. with excluded items (based on fit values and/or item discrimination). However, the studies also highlight some issues, such as test time constraints, the problems of open tasks, the problems of short scales with diverse items and possible, although not very pronounced ceiling effects. Furthermore, the scales often have low values of Cronbach's alpha. Moreover, the affective part of the model could not be tested in study 5. Additionally, study 5 does not include level 4 systemic thinking items.

The items partly showed difficulties that were different than expected. This could have several possible reasons, either individually or in combination:

(1) It could be a function of the items/instruments used, which are different than in the studies by Orion and his colleagues (see e.g. Ben-Zvi Assaraf & Orion, 2005a; 2005b; Ben-Zvi Assaraf & Orion, 2010b; Orion & Basis, 2008). Furthermore, the instruments/items of Orion and his colleagues are much more

detailed. Task format is one of the parameters that can influence item difficulty (see also e.g. Hammann, et al., 2008; Moosbrugger & Kelava, 2007, p. 249, with reference to Hartig & Klieme 2006, p. 136).

(2) The difficulty of certain sub-skills could be different based on the topic used, as neither of the studies of Orion and his colleagues uses 'tourism in Kenya'. In general, study 3 shows some differences in difficulty in the concept map items dealing with different sub-systems. Moreover, the InterGEO II showed differences in performance of German students on different sub-scales such as human and physical geography (Niemz & Stoltman, 1993) (see also the results of the study by LeVasseur, 1999). Similarly, Ben-Zvi Assaraf and Orion (2005a, p. 557) argue:

> "However, since research in education in the area of the earth systems is in a preliminary stage, more research is needed in order to test the current findings in relation to additional learning events, different age levels, different earth system subjects, and different cultures. Moreover, it is also suggested that the current findings and their interpretation should be tested in the context of other systems, namely technological, physical, biological, and sociological."

(3) It could also be a matter of analyses, as neither of Orion and his colleagues' papers (Ben-Zvi Assaraf & Orion, 2005a, b; Ben-Zvi Assaraf & Orion, 2010b; Orion & Basis, 2008) employs IRT to test the levels.

(4) Moreover, it could point to a difference between samples, for instance Israeli and German students, e.g. due to differences in the school system or culture and conceptions of learning (see also e.g. Ben-Zvi Assaraf & Orion, 2005a, p. 557; Lee, et al., 2007).

Overall, these results point to the need for further exploration. Regarding point (1), it would be useful to let items be rated by a number of experts in terms of their fit to a particular 'cell' of the competency model before employing them (e.g. Jo & Bednarz, 2008; Siegmund, et al., 2012). This could be done for instance using the delphi-method (see e.g. Horx Zukunftsinstitut GmbH, 2011). Moreover, it could be useful to develop items to create an instrument parallel to the Israeli one, and testing one sample with both to study the influences of the instrument/ type of items used on measured systemic thinking performance.

Point (2) could be explored by constructing parallel tests for two topics and/or regional examples, each covering the whole spectrum of systemic thinking skills and using as far as possible similar items (e.g. in terms of item format/length). Thereby, it would be useful to have comparative studies in pairs such as (a) same topic, different regional examples (e.g. 'tourism in country A' vs. 'tourism in country B'), (b) different topics within one broader thematic area, same regional example (e.g. 'tourism in country A' vs. 'agriculture in country A') and (c) topics from different broad thematic areas, same regional example (e.g. 'tourism in country A' vs. 'geology in country A'). Additionally, it would be helpful to explore the shares of the particular topic, general geographic systemic thinking skills and general problem solving skills in dealing with tasks requiring GSC in order to further develop the model. First results regarding the relationship between the facet model building of dynamic problem solving and geographic systemic thinking for university students show a latent correlation of $r=0.87$ ($p<0.001$) (HEIGIS project, Funke, et al., 2011; Siegmund, et al., 2012), which is high but comparable for instance to those of problem solving and maths ($r=0.89$), reading ($r=0.82$) or science ($r=0.80$) competencies in PISA 2003 (PISA-Konsortium Deutschland, 2004, p. 167).

Point (3) is difficult to address, as e.g. Ben-Zvi Assaraf and Orion's (2005a, b; Ben-Zvi Assaraf & Orion, 2010b) studies include combinations of diverse instruments such as Likert questions, drawing analyses, word association, concept mapping, interviews, a repertory grid and observation. Their description of the coding process points to the instruments not consisting of individual items coded as 0 or 1 as would be necessary for a Rasch model to be applied. Potentially, a study conducted with parallel instruments such as suggested with regard to point (1) also could employ IRT and an analyses done in the style of the original studies to gain some insight into the question.

Lastly, point (4) could be addressed by using the same instrument, in translated form, for both an Israeli and a German sample comparable in factors such as grade. Analyses could be done both overall and in terms of e.g. differential item functioning.

8.4.2 Spatial thinking

After the first study, it was decided to include spatial thinking as a variable. At first, the scales were too short and there was a need to simplify the graphics. In study 4, a Rasch scale was achieved but with a very pronounced ceiling effect. In study 5 some items were problematic and additionally, there still was a pronounced ceiling effect. Therefore, future studies should continue to develop spatial thinking scales covering the range from basic to more advanced skills. A useful framework for such a development could be e.g. the spatial thinking skills outlined by Gersmehl and Gersmehl (Gersmehl & Gersmehl, 2006, 2007). These were also the basis for the third dimension of the HEIGIS project (see e.g. Funke, et al., 2011; Siegmund, et al., 2012).

8.4.3 Model development

In general, the studies contributed to test instrument and model development for systemic thinking. Some further model and test instrument development has been done within the HEIGIS project for both systemic and spatial thinking (see e.g. Funke, et al., 2011; Siegmund, et al., 2012). Scale development for systemic thinking based on a different model is also conducted in a project by Rempfler and Uphues (e.g. 2012, see ch. 2). Furthermore, there are several ongoing projects dealing with aspects of systemic thinking in Germany which include topics relevant to the geosciences. For instance, Jahn's dissertation project (see e.g. Jahn & Siegmund, 2012) focuses on a study with high stream students to explore the effects of using satellite images vs. topographic maps on systemic thinking in the context of sustainable development. Moreover, the project "Shaping the future" examines kindergarten students' basic understanding and systemic thinking skills in the area of renewable energies, especially wind, water and solar power (see e.g. short project description on http://www.rgeo.de/cms/p/pzukunft/). Impulses for future development can also be received from projects such as "Space4Geography", which aims at developing a learning platform that allows students to explore a variety of topics with the help of remote sensing data and is planned to employ adaptive learning paths (Fuchsgruber, Wolf, Viehrig, Naumann, & Siegmund, 2014) or research accompanying the new Geo-ecological Laboratory at the Heidelberg University of Education (Volz & Sieg-

mund, 2013). Based on previous works, the results of these different studies, as well as those of current international ones, would need to be drawn together into one coherent model. Overall, however, there is still much work to be done both with regards to systemic and to spatial thinking scales.

For this, some of the open questions would relate to a further model specification. For instance, Fig. 71 shows two examples of possible network structures. Which does already count as exemplifying level 3 of systemic thinking, and in how far could the difference between the two be visible in the model? Similarly, both "water ---is heated by---> sun" and "water ---is heated due to absorbed---> global radiation" (see e.g. Forkel, 2012) are both connections between two concepts and actually describe the same process. Yet, while the first can be grasped already by young children, it is doubtful the second is as easy. However, the role of degrees of difficulty of elements or connections is till now not explicitly explained in the model.
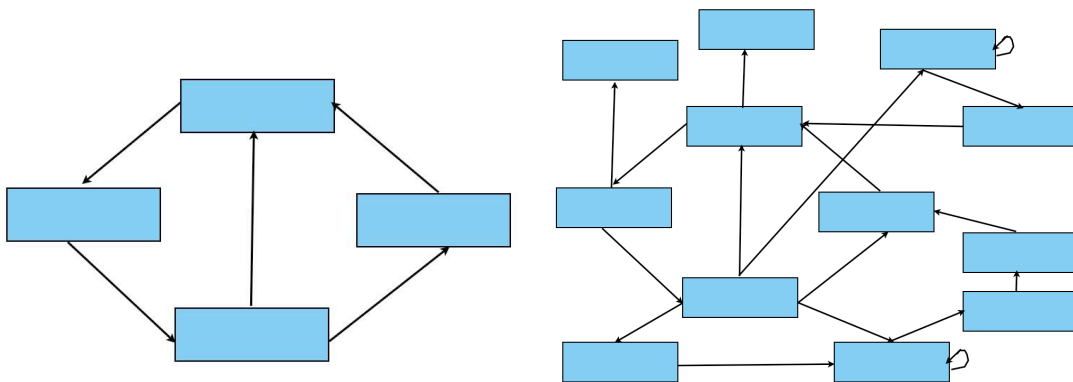
Fig. 71: Sample schemata of possible network structures

The studies covered a range of students from grade seven to university, with the focus on grade seven students. However, due to differences in the instruments used, for systemic thinking the individual studies cannot be combined. Helpful for the future would be studies employing for instance multiple-matrix-designs to study students' skills from beginner to advanced levels across a wide age range while at the same time reducing test time requirements for individual students (see e.g. Gonzalez & Rutkowski, 2010; and arguments in Kollar, 2012; Siegmund, et al., 2012).

At the end, geography education research and practice would benefit from a "Common Framework of References for Geography Education" similar to the "Common European Framework of References for Languages" (CEFR-L,

Council of Europe, 2001, n.d.) in the area of languages (see also e.g. Lenz, 2003, with reference to Vollmer). The framework should contain a variety of model descriptions (e.g. both those giving a general overview and those providing a more finely grained view), a set of guidelines for item construction for each level as well as a variety of sample exams and materials to aid in building instruments which often need to be specific to a particular sub-topic in order to test the effectiveness of particular interventions. As an example, the CEFR-L has competency descriptions on six levels (A1, A2, B1, B2, C1, C2) in several degrees of detailedness (overall, by sub-skill such as listening or reading, by detailed self-assessment statements) (Council of Europe, 2001, n.d.) and an accompanying document, the language passport, to document skills and experiences (EU, 2014). For a broad spectrum of languages, including e.g. Spanish and Russian, for instance textbooks (e.g. Adler & Bolgova, 2010; Lloret Ivorra, Ribas, & Wiener, 2009) and courses (e.g. Dialog Sprachreisen, 2014) reference the CEFR-L level to be achieved, and sample exams are available online (e.g. Instituto Cervantes, 2014; Universität Heidelberg, n.d.). Moreover, state educational curricula also often reference the CEFR-L levels (e.g. MKBW, 2004a) and the levels are recognized widely for specifying requirements in a broad range of areas (e.g. Auswärtiges Amt, 2014; RTVE.ES, 2014; Ruslanguage School Moscow, 2014; Universität Erfurt, 2014; Universität Heidelberg, 2014; Wimdu GmbH, 2014).

Such model descriptions and scales making self-assessment possible are becoming increasingly important. For instance, in Baden-Wurttemberg, recent educational policy changes include the introduction of the so called *Gemeinschaftsschule*, in which students aiming at high, mid, low stream as well as special needs certificates should study together (MKBW, 2012). This also entails developing competency rubrics, check lists, individual learning tasks and forms of planning and reflecting the students' individual work (e.g. LfS, 2012; LfS, n.d.; MKBW & LfS, n.d.; Müller, 2003). Notably, the Swiss private school *Institut Beatenberg,* which is one of the models for the policies developing the *Gemeinschaftschule* (see e.g. Müller, 2013), has published competency rubrics for a variety of subjects including geography (as part of "world"), earth science (as part of "science") and computer science, using the level names (A1 etc.) of

the language framework (Institut Beatenberg, n.d.). Additionally, in learning theory, self-assessment and reflexion of one's own pre-knowledge and learning process are seen as helpful for successful learning (e.g. Benson, 2001; Bimmel & Rampillion, 2000; Meyer, 2006; more critically, Schmenk, 2004).

8.4.4 Methodology

The studies also highlighted some advantages and disadvantages of using paper and pencil test instruments. Especially for a larger sample, such as was achieved in study 5, such instruments mean a substantial amount of time needed to enter the data. While paper and pencil tests have the organizational advantage of not having to go to the computer lab, they also mean having to receive, hand out, collect, and send back a stack of paper. Computer-based instruments also have a broader spectrum of utilization, for instance in terms of a subsequent inclusion in learning platforms. In general, paper based concept maps in a pre-/posttest setting can be influenced by motivational issues (Fiene, 2014).

One possibility for creating computer-based instruments is Lime-Survey (http://www.limesurvey.org/en/), which is used e.g. by the HEIGIS project and the study by Kollar (2012). Yet, computer-based instruments also have limitations e.g. with regard to item formats. For instance, Lime-Survey does not (yet) allow for an own creation of concept maps. Other software to create and analyze concept maps on the computer exists, such as MaNet (see e.g. Stracke, 2004) or CMapTools (http://cmap.ihmc.us/). Overall, the most advantageous would be questions that can be solved both in paper and pencil and computer-based form in a compact format.

In general, there is still much work to be done to find the most favorable item formats as well as clearly understand factors associated with item difficulty beyond the competency to be tested. Thereby, results from other domains e.g. regarding the length and structure of the item stem text could be a useful starting point (see e.g. overview in Freunberger & Itzlinger-Bruneforth, 2013).

8.5 Overall outlook

"The field of geography education is sadly lacking in empirical data that might inform and underpin decisions [...]" wrote Downs two decades ago (1994, p.

57). For the German situation, Lethmate (2001, p. 40, translated) criticized that results from other research areas are not taken into account, and states that "[g]eography, it seems, happens in empirical vacuum." Much research has been conducted since then, to which the present studies are a small contribution. Yet even now, the call for high-quality, cumulative research is still relevant, as so far, there are still too many little understood spots (see also e.g. Baker, et al., 2012; Bednarz, et al., 2013). As Bednarz et al. (2013, p. 7) summarize:

> "We need better and more research before we can understand even the
> most fundamental ways individuals develop proficiency in geography."

This also includes "[...] the need to clarify what it means to "do geography"" (ibid, p. 23). In contrast to the area of languages, there is still no "Common Framework of Reference for the (Geo)sciences" which would describe internationally recognized levels, help students evaluate their own achievement independently of the source of their competencies (autonomous learning, courses in school, etc.) and underpin research into and teaching for competency improvement.

Overall, the studies have not given support for the suitability of a fairly low cost version of GIS (WebGIS, working with a parter, pre-fabricated materials the students can work themselves with, short duration) for fostering systemic thinking. Due to the issues with the test instrument, sample, and treatment, however, these need to be further examined. Especially, it needs to be falsified that the negative result is not only due to students' lack of GIS pre-experience and thus not being able to concentrate fully on the geographic system. It has to be kept in mind that it is difficult to examine the effects of working with one medium in general. As Genevois and Joliveau argue (2009, p. 116):

> "As for methodology, we chose not to make comparisons between
> test groups and control groups: a successful (or failed) experiment
> under specific conditions may give very different results if we change
> a variable. The complexity of the classroom and the many didactic
> variables make the comparison of representative samples illusory"

Despite this, to explore variables associated with competency development, especially in central areas, seems necessary in order to improve education in the geo-sciences.

9. References

"Binary data and factor analysis". (2000-2011). responses on 2007-11-07 5:55 pm by Linda K. Muthen, 2007-11-07 6:50 by Jon Elhai, 2007-11-08 4:42 by Bengt O. Muthen, 2008-01-31 10:46 am by Linda K. Muthen. Retrieved 2011-09-06, from http://www.statmodel.com/discussion/messages/8/50.html

Aarons, M. (2003). *GIS as effective tools in geography education. Their influence on raising pupils' geographical attainment.* Master of Arts, University of London, London.

Ackerman, E. A. (1963). Where is a research frontier? *Annals of the Association of American Geographers, 53*(4), 429-440.

Adler, I., & Bolgova, L. (2010). *Мост - Russisch für Anfänger. Lehrbuch mit 2 Audio-CDs*: Klett.

Åhlberg, M., & Ahoranta, V. (2002). Two Improved Educational Theory Based Tools to Monitor and Promote Quality of Geographical Education and Learning. *International Research in Geographical and Enviromental Education, 11*(2), 119-137.

Ahmed, F., & Capretz, L. F. (2008). The software product line architecture: An empirical investigation of key process activities. *Information and Software Technology, 50*, 1098-1113.

Akama, J. S. (1997). Tourism Development in Kenya: Problems and Policy Alternatives. *Progress in Tourism and Hospitality Research, 3*, 95-105.

Akama, J. S. (1999). The Evolution of Tourism in Kenya. *Journal of Sustainable Tourism, 7*(1), 6-25.

Akama, J. S. (2000). The Efficacy of Tourism as a Tool For Economic Development in Kenya. *DPMN Bulletin Tourism and African Development, VII*(1).

Akama, J. S. (2002). The Creation of the Maasai image and tourism development in Kenya. In J. S. Akama & P. Sterry (Eds.), *Cultural tourism in Africa: strategies for the new millennium, Proceedings of the ATLAS Africa International Conference, December 2000, Mombasa, Kenya* (pp. 43-53). Arnhem: ATLAS.

Aladag, E. (2007). Iköğretim 7. Sınıf Sosyal Bilgiler Dersinde Coğrafi Bilgi Sistemleri Kullanımıınn Öğrencilerin Akademik Başari Ve Derse Karşi Motivasyonlarina Etkisi. Retrieved 2014-05-02, from http://acikarsiv.gazi.edu.tr/index.php?menu=2&secim=10& YayinBIK=293

Aladag, E. (2009). The Opinions of Turkish Social Studies Pre-Service Teachers about Using GIS in Primary Education. In T. Jekel, A. Koller & K. Donert (Eds.), *Learning with Geoinformation IV - Lernen mit Geoinformation IV* (pp. 39-49). Heidelberg: Wichmann.

Aladag, E., & Aladag, S. (2008). The use of GIS in primary schools in Turkey. In A. Demirci, M. Karakuyu, M. A. McAdams, S. Incekara & A. Karaburun (Eds.), *5th International Conference on Geographic Information Systems. Proceedings* (pp. 171-177). Istanbul: Fatih University Department of Geography.

Albrecht, K. (1998). Der Beitrag des Tourismus zur wirtschaftlichen Entwicklung Kenias. Darstellungsschwerpunkt: Arbeit mit Statistiken im Erdkundeunterricht einer 9.Klasse eines Berliner Gymnasiums. Retrieved 2007-12-11, from http://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/zielgruppen/lehramtsanwaerterinnen/erdkunde/Katja_Albrecht__DerBeitragdes_TourismuszurwirtschaftlichenEntwicklungKenias1.pdf

Alibrandi, M. (2003). *GIS in the classroom: using geographic information systems in social studies and environmental science*. Portsmouth: Heinemann.

Arndt, H. (2006). Modellierung und Simulation im Wirtschaftsunterricht zur Förderung systemischen und prozessorientierten Denkens am Beispiel unternehmensübergreifender Kooperation in Wertschöpfungsketten. *bwp*(10).

Audet, R., & Abegg, G. L. (1996). Geographic information systems: Implications for problem solving. *Journal of Research in Science Teaching, 33*(1), 121-145.

Audet, R. H., & Paris, J. (1997). GIS implementation model for schools: Assessing the critical concerns. *Journal of Geography, 96*(6), 293-300.

Auswärtiges Amt. (2014). Fremdsprachenassistentinnen/ Fremdsprachenassistenten für Englisch in Kombination mit Chinesisch, Portugiesisch oder Russisch. Retrieved 2014-06-23, from http://www.stepstone.de/stellenangebote--Fremdsprachenassistentinnen-Fremdsprachenassistenten-fuer-Englisch-Berlin-Auswaertiges-Amt--2906156-inline.html?cid=msearche_indeed___C

Backes, M. (2005). Das ganze Land in einem Dorf "Die Bomas of Kenya". *Überblick, 11*(2), 6-11.

Baenninger, M., & Newcombe, N. (1989). The Role of Experience in Spatial Test Performance: A Meta-Analysis. *Sex Roles, 20*(5/6).

Baker, T. R. (2002). *The effects of Geographic Information System (GIS) technologies on students' attitudes, self-efficacy, and achievement in middle school science classrooms.* Dissertation, University of Kansas.

Baker, T. R. (2005). Internet-Based GIS Mapping in Support of K-12 Education. *The Professional Geographer, 57*(1), 44-50.

Baker, T. R., & Bednarz, S. W. (2003). Lessons learned from Reviewing Research in GIS education. *Journal of Geography*(102), 231-233.

Baker, T. R., Kerski, J. J., Huynh, N. T., Viehrig, K., & Bednarz, S. W. (2012). Call for an Agenda and Center for GIS Education Research. *RIGEO, 2*(3), 254-288.

Baker, T. R., Palmer, A. M., & Kerski, J. J. (2009). A National Survey to Examine Teacher Professional Development and Implementation of Desktop GIS. *Journal of Geography, 108*(4-5), 174-185.

Baker, T. R., & White, S. H. (2003). The Effects of GIS on Students' Attitudes, Self-efficacy, and Achievement in Middle School Science Classrooms. *Journal of Geography*(102), 243-254.

Ballantyne, R. (1999). Teaching Environmental Concepts, Attitudes and Behaviour Through Geography Education: Findings of an International Survey. *International Research in Geographical and Environmental Education, 8*(1), 40-58.

Barke, H.-D., & Engida, T. (2001). Structural chemistry and spatial ability in different cultures. *Chemistry Education: Research and Practice in Europe, 2*(3), 227-239.

Battersby, S. E., Golledge, R. G., & Marsh, M. J. (2006). Incidental Learning of Geospatial Concepts Across Grade Levels: Map Overlay. *Journal of Geography, 105*, 139-146.

Bauer, J., Englert, W., Meier, U., Morgeneyer, F., & Waldeck, W. (2002). *Physische Geographie kompakt*. Heidelberg Berlin: Spektrum Akad. Verl.

Bauer, J., Hallermann, S., Morgeneyer, F., Schreiner, A., Schwarz, A., & Waldeck, W. (2007). *Aspekte der Globalisierung - ein Methodenband -*. Braunschweig: Schroedel.

Baumann, M., Colditz, M., Irmscher, U., Gerber, W., Protze, N., Reutemann, S., et al. (2004). *Heimat und Welt 7. Mittelschule Sachsen*. Braunschweig: Westermann.

Bayrhuber, H., Häußler, P., Hemmer, I., Hemmer, M., Hlawatsch, S., Hoffmann, L., et al. (2002). Interesse an geowissenschaftlichen Themen. Ergebnisse einer Interessenstudie im Rahmen des Projekts „Forschungsdialog System Erde". *Geographie heute*(202), 22-23.

BBN Corporation. (1997a). Prophet StatGuide: Do your data violate one-way ANOVA assumptions? Retrieved 2013-01-06, from http://www.basic.northwestern.edu/statguidefiles/oneway_anova_ass_viol.html

BBN Corporation. (1997b). Prophet StatGuide: Do your data violate t test assumptions? Retrieved 2013-01-06, from http://www.basic.northwestern.edu/statguidefiles/ttest_unpaired_ass_viol.html

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A., & Kelly, D. (1996). *Science Achievement in the middle school years: IEA's third international mathematics and science study (TIMSS)*: IEA.

Beauftragte der Bundesregierung für Migration Flüchtlinge und Integration (Ed.). (2005). *Daten-Fakten-Trends. Bildung und Ausbildung*. Berlin.

Bednarz, R. S., & Bednarz, S. W. (2008). The effect of explicit instruction in spatial thinking. In A. Demirci, M. Karakuyu, M. A. McAdams, S. Incekara & A. Karaburun (Eds.), *5th International conference on geographic information systems: proceedings* (pp. 649-655). Istanbul: Fatih University Department of Geography.

Bednarz, S. W. (2001). Thinking Spatially: Incorporating Geographic Information Science in Pre and Post Secondary Education. Retrieved 2007-12-17, from www.geography.org.uk/projects/spatiallyspeaking/furthermaterials/

Bednarz, S. W., Heffron, S., & Huynh, N. T. (Eds.). (2013). *A Road Map for 21st Century Geography Education. Geography Education Research. Recommendations and Guidelines for Research in Geography Education*. Washington D.C.: Association of American Geographers.

Bell, T. (2004). Komplexe Systeme und Strukturprinzipien der Selbstregulation – Konstruktion grafischer Darstellungen, Transfer und systemisches Denken. *Zeitschrift für Didaktik der Naturwissenschaften, 10*, 183-204.

Ben-Zvi Assaraf, O., & Orion, N. (2005a). Development of System Thinking Skills in the Context of Earth System Education. *Journal of Research in Science Teaching, 42*(5), 518-560.

Ben-Zvi Assaraf, O., & Orion, N. (2005b). A Study of Junior High Students' Perceptions of the Water Cycle. *Journal of Geoscience Education, 53*(4), 366-373.

Ben-Zvi Assaraf, O., & Orion, N. (2010a). Four case studies, six years later: developing system thinking skills in Junior High School and sustaining them over time. *Journal of Research in Science Teaching, 47*(10), 1253-1280.

Ben-Zvi Assaraf, O., & Orion, N. (2010b). System thinking skills at the elementary school level. *Journal of Research in Science Teaching, 47*(5), 540-563.

Benedikt, J., & Danhofer, M. (2005). Soziodemografische Analysen mit dem ArcExplorer, Visualisierung von Segregationsprozessen in nordamerikanischen Städten am Beispiel New York City. *Geographie heute, 26*(233), 26-29.

Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow Longman.

Bertelsmann Stiftung. (n.d.). Folgekosten unzureichender Bildung. Retrieved 2010-03-23, from http://www.bertelsmann-stiftung.de/cps/rde/xchg/SID-7BED654B-FEF7C7D5/bst/hs.xsl/93522_93536.htm

Betts, J. R. (2004). Peer Groups and Academic Achievement: Panel Evidence from Administrative Data. *IZA Paper*. Retrieved from http://www.iza.org/iza/en/webcontent/events/transatlantic/papers_2004/betts.pdf

Bialystok, E. (2008). Second-Language Acquisition and Bilingualism at an Early Age and the Impact on Early Cognitive Development. In R. E. Tremblay, R. G. Barr & R. D. Peters (Eds.), *Encyclopedia on Early Childhood Development (online)*. Montreal: Centre of Excellence for Early Childhood Development.

Bimmel, P., & Rampillion, U. (2000). *Lernerautonomie und Lernstrategien*. München.

Black, A. A. J. (2005). Spatial Ability and Earth Science Conceptual Understanding. *Journal of Geoscience Education, 53*(4), 402-414.

BLK (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung). (1999). *Bildung für eine nachhaltige Entwicklung*. Bonn.

BMBF. (2008). 25 Jahre Leben in Deutschland – 25 Jahre Sozio-oekonomisches Panel. Retrieved 2010-03-23, from http://www.bmbf.de/pub/soep_leben_in_deutschland.pdf

Board on Earth Sciences and Resources. (2006). *Beyond mapping: meeting national needs through enhanced geographic information science*. Washington: National Academies Press.

Bodzin, A. M. (2008). Integrating Instructional Technologies in a Local Watershed Investigation With Urban Elementary Learners. *The Journal of Environmental Education, 39*(2), 47-57.

Bodzin, A. M., & Anastasio, D. (2006). Using Web-based GIS For Earth and Environmental Systems Education. *Journal of Geoscience Education, 54*(3), 297-300.

Bollmann-Zuberbühler, B. (2008). Lernwirksamkeitsstudie zum systemischen Denken an der Sekundarstufe I. In U. Frischknecht-Tobler, U. Nagel & H. Seybold (Eds.), *Systemdenken. Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (pp. 99-118). Zürich.

Bond, & Fox. (2007). *Applying the Rasch Model.: Fundamental Measurement in the Human Sciences*. Mahwah: Lawrence Erlbaum Associates.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.

Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., et al. (Eds.). (2008). *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich. Zusammenfassung*. Berlin.

Bradshaw, M., & Stratford, E. (2000). Qualitative Research Design and Rigour. In I. Hay (Ed.), *Qualitative research methods in human geography* (pp. 37-49). Oxford.

Branch, R. M., & Gustafson, K. L. (1998). Re-visioning models of instructional development. *Paper presented at the Annual Meeting of the Association for Educational Communications and Technology*. Retrieved from http://ocw.metu.edu.tr/file.php/118/Week7/gustafson_id-models-ED416837_3.pdf

Briebach, T. (2007). What impact has GIS had on geographical education in secondary schools? Retrieved 2008-01-29 from http://www.geography.org.uk/projects/spatially speaking/participantreports/

Brind, T., Harper, C., & Moore, K. (2008). Education for migrant, minority and marginalised children in Europe. A report commissioned by the Open Society Institute's Education Support Programme. Retrieved 2010-04-14, from http://www.soros.org/initiatives/esp/articles_publications/publications/children_20080131/review_20080131.pdf

Brodie, S. (2004). *GIS in years 7 - 13 social science education: a New Zealand perspective*. presented at the ESRI Education User Conference 2004. Retrieved from http://gis.esri.com/library/userconf/educ04/papers/pap5097.pdf

Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.

Bullinger, R., & Hieber, U. (2004). "Werte das Klimadiagramm aus!" Klimadiagramme in der schriftlichen Lernzielkontrolle. *Geographie heute*(224), 16-21.

Bundesamt für Kartographie und Geodäsie. (2003). *Geoinformation und moderner Staat*. Frankfurt/Main.

Bundeszentrale für politische Bildung. (2008). ecopolicyade. Retrieved 2008-05-06, from www.bpb.de/veranstaltungen/QDBEXE,0,ecopolicyade.html

Burgmaier, F., & Traub, A. (2007). Schüler mit Migrationshintergrund. Auf die Definition kommt es an! *Zeitischrift für Bildungsverwaltung, 2*(07), 5-16.

Burke, M. A., & Sass, T. R. (2008). Peer effects and student achievement Available from http://www.urban.org/UploadedPDF/1001190_peer_effects.pdf?RSSFeed=UI_Education.xml

Buske, H. G. (2004). Durch eigene Erfahrung lernen. Methodenerwerb im Erdkundeunterricht. *Geographie und Schule*(149), 2-8.

Butt, G., & Weeden, P. (2004). Boys' Underachievement in Geography: An Issue of Ability, Attitude or Assessment? *International Research in Geographical and Environmental Education, 13*(4), 329-347.

Campbell, D. J., Gichohi, H., Reid, R., Mwangi, A., Chege, L., & Sawin, T. (2003). Interactions between People and Wildlife in Southeast Kajiado District, Kenya. *LUCID Working Paper Series*(18).

Carneiro, R. (2006). Motivating School Teachers to Learn: can ICT add value? *European Journal of Education, 41*(3/4), 415-435.

Clagett, K. E. (2009). *Virtual Globes as a platform for developing spatial literacy.* Universidade Nova de Lisboa Lisbon.

Clark, A. M., Monk, J., & Yool, S. R. (2007). GIS pedagogy, web-based learning and student achievement. *Journal of Geography in Higher Education, 31*(2), 225-239.

Clough, & Holden. (2002). Tourism: do we want the new hotel? (pp. 122-126). Retrieved 2010-08-24

Cohen, J. (1988). 2.2.3 "SMALL", "MEDIUM," AND "LARGE" d VALUES. In J. n. e. Cohen (Ed.), *Statistical power analysis for the behavioral sciences* (pp. 24-27): Routledge.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5 ed.). London Routledge Falmer.

College Board. (2007). AP Human Geography Teacher's Guide. Retrieved 2010-03-22, from http://apcentral.collegeboard.com/apc/members/repository/ap07_humangeo_teachers_guide.pdf

Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology, 78*(1), 98-104.

Council of Europe. (2001). Common European Framework of Reference For Languages. Retrieved 2007-10-22, from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Council of Europe. (2002). European Charter for Regional or Minority Languages. Application of the Charter in Germany. Retrieved 2010-07-05, from http://www.coe.int/t/dg4/education/minlang/report/EvaluationReports/GermanyECRML1_en.pdf

Council of Europe. (2010). List of declarations made with respect to treaty No. 148, status as of: 5/7/2010. Retrieved 2010-07-05, from http://conventions.coe.int/treaty/Commun/ListeDeclarations.asp?NT=148&CM=1&DF=&CL=ENG&VL=1

Council of Europe. (n.d.). Overview of CEFR-related scales. Retrieved 2014-06-23, from http://www.coe.int/t/dg4/education/elp/elp-reg/cefr_scale_EN.asp

cpa/dpa/AP. (2006). Zehntklässler-Pisa: Jeder dritte Schüler lernt nichts hinzu. Retrieved 2010-03-22, from http://www.spiegel.de/schulspiegel/wissen/0,1518,449011,00.html

Crechiolo, A. L. (1997). *Teaching secondary school geography with the use of a geographical information system (GIS)* MA thesis, Wilfrid Laurier University.

Cremer, P., Richter, B., & Schäfer, D. (2004). GIS im Geographieunterricht – Einführung und Überblick. *Praxis Geographie, 34*(2), 4-7.

Crews, J. W. (2008). *Impacts of a teacher geospatial technologies professional development project on student spatial literacy skills and interests in science and technology in grades 5-12 classrooms across Montana.* Dissertation, University of Montana, Missoula.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronin, M., Gonzales, C., & Sterman, J. D. (2009). Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens. *Organizational Behavior and Human Decision Processes 108*(1), 116-130.

Dallal, G. E. (2000). Nonparametric Statistics. Retrieved 2011-08-05, from http://www.JerryDallal.com/LHSP/npar.htm

Dallal, G. E. (2003). Why p=0.05? Retrieved 2011-05-05, from http://www.JerryDallal.com/LHSP/p05.htm

Davidz, H. L., Nightingale, D. J., & Rhodes, D. H. (2004). Enablers, Barriers, and Precursors to Systems Thinking Development: The Urgent Need for More Information. Retrieved 2016-06-30, from http://esd.mit.edu/HeadLine/pdfs/davidz1_incose_092004.pdf

Dawson, G. (n.d.). Year 8 Virtual Holiday to Kenya Assessment. Retrieved 2010-08-24, from http://www.sln.org.uk/geography/fairoak/documents/Year_8_Tourism_in_Kenya_Assessment.doc

de Jong, T. (2010). Cognitive load tehory, educational research, and instructional design: some food for thought. *Instructional Science, 38*(2), 105-134.

de Lange, N. (2005). *Geoinformatik in Theorie und Praxis*. Berlin.

de Lange, N. (2006a). *Geoinformationssysteme im Geographieunterricht: Paradigmenwechsel?!* presented at the Vortrag auf der 2. GIS-Ausbildungstagung 2006 Potsdam, Konferenz-CD.

de Lange, N. (2006b). Geoinformationssysteme in Schulen: derzeitiger Stand und zukünftiger Einsatz. In T. Jekel, A. Koller & J. Strobl (Eds.), *Lernen mit Geoinformation* (pp. 11-22). Heidelberg: Wichmann.

de Lange, N. (2007). GIS in der Geoinformatik - GIS in der Schule. In T. Jekel, A. Koller & J. Strobl (Eds.), *Lernen mit Geoinformation II* (pp. 32-41). Heidelberg: Wichmann.

Demirci, A. (2008). Evaluating the Implementation and Effectiveness of GIS-Based Application in Secondary School Geography Lessons. *American Journal of Applied Sciences, 5*(3), 169-178.

DGfG. (2002). *Arbeitsgruppe Curriculum 2000+. Grundsätze und Empfehlungen für die Lehrplanarbeit im Schulfach Geographie*. Bonn.

DGfG. (2007a). *Bildungsstandards im Fach Geographie für den Mittleren Bildungsabschluss*. Berlin.

DGfG. (2007b). *Educational Standards in Geography for the Intermediate School Certificate*. Berlin.

DGfG. (2008). *Bildungsstandards im Fach Geographie für den Mittleren Bildungsabschluss*. Berlin.

DGfG (Ed.). (2010). *Rahmenvorgaben für die Lehrerausbildung im Fach Geographie an deutschen Universitäten und Hochschulen*. Bonn: Selbstverlag Deutsche Gesellschaft für Geographie.

Dialog Sprachreisen. (2014). Häufige Fragen rund um das Thema Sprachreisen mit DIALOG Retrieved 2014-06-23, from http://www.dialog.de/service/haeufige-fragen-faq/#c5142

Dickerson, D., & Dawkins, K. (2004). Eighth Grade Students' Understandings of Groundwater. *Journal of Geoscience Education, 52*(2), 178-181.

Diezmann, C. M., & Watters, J. (2000). Identifying and supporting spatial intelligence in young children. *Contemporary Issues in Early Childhood, 1*(3), 299-313.

Dissen, H. (1990). Tourismus und Entwicklung in der Dritten Welt am Beispiel von Kenia. *Geographie und Schule, 12*(64), 35-38.

Ditter, R. (2013). *Die Wirksamkeit digitaler Lernwege in der Fernerkundung. Eine empirische Untersuchung zu Lernmotivation und Selbstkonzept bei Schülerinnen und Schülern der Sekundarstufe.* Heidelberg University of Education, Heidelberg.

Ditton, H. (1998). *Mehrebenenanalyse. Grundlage und Anwendungen des Hierarchisch Linearen Modells*. Weinheim: Juventa.

Doering, A., & Veletsianos, G. (2008). An Investigation of the Use of Real-Time, Authentic Geospatial Data in the K–12 Classroom. *Journal of Geography, 106*(6), 217-225.

Donert, K. (2007). Aspects of the State of Geography in European higher education. TUNING Geography: a report of findings and outcomes. *HERODOT*. Retrieved from http://www.herodot.net/state/TUNING-Geography-v1.pdf

Downs, R. (1994). The Need for Research in Geography Education: It Would be Nice to Have Some Data. *Journal of Geography, 93*(1), 57-60.

Drennon, C. (2005). Teaching Geographic Information Systems in a Problem-Based Learning Environment. *Journal of Geography in Higher Education, 29*(3), 385-402.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education 25*(6), 671-688.

Duke, B. A., & Kolvoord, B. (n.d.). *STEM. National Center for Rural Science, Technology, Engineering & Mathematics Education Outreach. Geospatial Technology. Curriculum Page*.

Durpoix, D. (2010). The factor analysis techniques used in the study *Farmers' attitudes and behaviour towards the natural environment: a New Zealand case study* (pp. 333-334). Massey University, Palmerston North, New Zealand: Thesis.

Edelmann, W. (2000). *Lernpsychologie*. Weinheim: Beltz.

Einhaus, E. (2007). *Schülerkompetenzen im Bereich Wärmelehre. Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*. Berlin: Logos.

Ellis, P. D. (2009). Effect size calculators. Retrieved 2014-06-16, from http://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html

Engelmann, D., Faust-Ern, U., Gaffga, P., Kreuzberger, N., Latz, W., Püschel, L., et al. (2005). *Diercke Erdkunde für Gymnasien in Nordrhein-Westfalen 8*. Braunschweig: Westermann.

Erhard, A. (2003). Tourismus und sozioökonomische Wandel: das Beispiel der Insel Lamu (Kenia). *Geographische Rundschau, 55*(7-8), 18-23.

Erpenbeck, J., & von Rosenstiel, L. (2003). Einführung. In J. Erpenbeck & L. von Rosenstiel (Eds.), *Handbuch Kompetenzmessung - Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (pp. IX-XL). Stuttgart: Schäffer-Poeschel.

Escamilla, K., Chávez, L., & Vigil, P. (2005). Rethinking the "Gap": High-Stakes Testing and Spanish-Speaking Students in Colorado. *Journal of Teacher Education, 56*(2), 132-144.

Esikuri, E. E. (1998). Spatio-Temporal Effects of Land Use Changes in A Savanna Wildlife Area of Kenya. [Dissertation].

ESRI. (1998). GIS in K-12 education. An ESRI White Paper. Retrieved 2007-04-10, from www.esri.com/industries/k-12/download/docs/k12educ2.pdf

ESRI. (2003). *Geographic Inquiry: Thinking Geographically*: ESRI Schools and Libraries Program.

ESRI. (2010). ArcGIS in Secondary Schools - Newsletter 01/2010. Retrieved 2010-04-08, from http://esri-rwanda.com/downloads/papers/ESRI_Newsletter_Ruanda_012010.pdf

ESRI. (n.d.). Company History. Retrieved 2008-11-29, from http://www.esri.com/company/about/history.html

EU. (2014). Europass. Retrieved 2014-07-01, from https://europass.cedefop.europa.eu/de/home

European Commission. (2006). Use of Computers and the Internet in Schools in Europe 2006. Country Brief: Germany. Retrieved 2007-07-11, from http://d21.fujitsu-siemens.de/12_06/images/CountryBrief_Germany.pdf

European Commission. (n.d.). About INSPIRE. Retrieved 2014-02-14, from http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48

Ewing, M. T., Salzberger, T., & Sinkovics, R. R. (2009). Confirmatory factor analysis vs. Rasch approaches: Differences and Measurement Implications. *Rasch Measurement Transactions,* (23:1), 1194-1195. Retrieved from http://www.rasch.org/rmt/rmt231g.htm

Example Calculation of Eta-squared from SPSS Output. Retrieved 2012-01-10, from http://wilderdom.com/research/ExampleCalculationOfEta-SquaredFromSPSSOutput.pdf

Falk, G. (2003). *Didaktik des computerunterstützten Lehrens und Lernens illustriert an Beispielen aus der geographieunterrichtlichen Praxis*. Berlin: Mensch & Buch.

Falk, G. (2004). Das Diercke GIS im projektorientierten Unterricht. In Y. Schleicher (Ed.), *Computer, Internet & Co. im Erdkundeunterricht* (pp. 191-210). Berlin: Cornelsen Scriptor.

Falk, G., & Hoppe, W. (2004). GIS – ein Gewinn für den Geographieunterricht? Überlegungen zum Einsatz moderner Geoinformationssoftware im Unterricht. *Praxis Geographie, 34*(2), 10-12.

Falk, G., & Nöthen, E. (2004). LÄRM! Schüler erforschen mit GIS stadtökologische Probleme. *Praxis Geographie, 34*(2), 35-38.

Falk, G., & Nöthen, E. (2005). *GIS in der Schule Potenziale und Grenzen*. Berlin: Mensch & Buch-Verlag.

Falk, G., & Schleicher, Y. (2005). Didaktik und Methodik des schulischen GIS-Einsatzes. *Geographie heute, 26*(233), 2-7.

Fashola, O. S., Slavin, R. E., Calderón, M., & Durán, R. (1997). *Effective Programs for Latino Students in Elementary and Middle Schools*: Center for Research on the Education of Students Placed At Risk (CRESPAR).

Favier, T., & Van der Schee, J. (2009). Learning to think geographically by working with GIS. In T. Jekel, A. Koller & K. Donert (Eds.), *Learning with Geoinformation IV - Lernen mit Geoinformation IV* (pp. 135-145). Heidelberg: Wichmann.

Federal Research Division. (2007). Country Profile: Kenya. Retrieved 2008-03-14, from http://lcweb2.loc.gov/frd/cs/profiles/Kenya.pdf

Ferguson, C. J. (2009). An Effect Size Primer: A Guide for Clinicans and Researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538.

Ferguson, R. F. (2002). What doesn't meet the eye: understanding and addressing racial disparities in high-achieving suburban schools. Retrieved 2009-06-15, from http://www.ncrel.org/gap/ferg/

Feyk, M. (2006). Kritische Gedanken zum Einsatz von Geografischen Informationssystemen (GIS) in allgemeinbildenden Schulen. *Medien + Erziehung, 50*(4), 63-69.

Fiene, C. (2014). *Wahrnehmung von Risiken aus dem globalen Klimawandel – eine empirische Untersuchung in der Sekundarstufe I.* Dissertation, Heidelberg University of Education.

Fitzpatrick, C., & Dailey, G. (2000). Learning with GIS. Retrieved 2008-11-29, from http://www.esri.com/news/arcuser/0700/umbrella11.html

Flath, M. (1994). Die Erde bewahren - Leitbild für den Geographieunterricht? In M. Flath & G. Fuchs (Eds.), *Die Erde bewahren - Fremdartigkeit verstehen und respektieren, Erstes Gothaer Forum zum Geographieunterricht, 26./27. November 1993 in Gotha* (pp. 24-28). Gotha: Justus Perthes Verlag.

Flath, M., & Fuchs, G. (1994). Thema 1: Lehrplanakzente: Die Erde bewahren - Leitidee für den Geographieunterricht?, Diskussionsprotokoll (strukturierte Zusammenfassung), *Die Erde bewahren - Fremdartigkeit verstehen und respektieren, Erstes Gothaer Forum zum Geographieunterricht, 26./27. November 1993 in Gotha* (pp. 29-33). Gotha.

Flath, M., & Kulke, E. (Eds.). (2007). *Mensch und Raum - Oberstufe Geographie*. Berlin: Cornelsen.

Flecke, S. (2001). *Entwicklung eines Virtuellen Schulbuches als interaktive, multimediale Lernumgebung.* Diplom-Thesis, Technische Universität, Darmstadt.

Forkel, M. (2012). Infoblatt Strahlungshaushalt. Einführung in den Strahlungshaushalt der Erde. Retrieved 2014-06-23, from http://www2.klett.de/sixcms/list.php?page=geo_infothek&article=Infoblatt+Strahlungshaushalt&node=Grundlagen+der+Klimatologie

Forster, M. (2008). Teaching GIS at Rwandan Secondary Schools. *ArcNews Online,* (Spring 2008). Retrieved from http://www.esri.com/news/arcnews/spring08articles/teaching-gis.html

Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeo, D. A., & Walker, D. A. (1962). *Educational achievements of thirteen-year-olds in twelve countries*. Hamburg: Unesco institute for education.

Frank, L., Maclennan, S., Hazzah, L., Bonham, R., & Hill, T. (2006). Lion Killing in the Amboseli-Tsavo Ecosystem, 2001-2006, and its Implications for Kenya's Lion Population. Retrieved 2014-05-02, from http://www.lionconservation.org/AnnualReports/2006-Lion-killing-in-Amboseli-Tsavo-ecosystem.pdf

Freksa, C., & Röhring, R. (1993). *Dimensions of qualitative spatial reasoning.* Paper presented at the Qualitative reasoning in decision technologies, QUARDET 1993, Barcelona.

Freunberger, R., & Itzlinger-Bruneforth, U. (2013). Item Entwicklung: Kognitive Prozesse und Item Schwierigkeit. Retrieved 2014-06-25, from https://www.bifie.at/system/files/dl/Item-Entwicklung.pdf

Frischknecht-Tobler, U., Kunz, P., & Nagel, U. (2008). Systemdenken - Begriffe, Konzepte und Definitionen. In U. Frischknecht-Tobler, U. Nagel & H. Seybold (Eds.), *Systemdenken. Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (pp. 11-31). Zürich: Pestalozzianum.

Frischknecht-Tobler, U., Nagel, U., & Seybold, H. (2008). *Systemdenken. Wie Kinder und Jugendliche komplexe Systeme verstehen lernen*. Zürich: Pestalozzianum.

Fuchsgruber, V., Wolf, N., Viehrig, K., Naumann, S., & Siegmund, A. (2014). Understanding the earth with satellite images - Development of a student-centered learning environment to support the application of remote sensing in schools. In B. Zagajewski, A. Sabat, M. Golenia, A. Robak, A. Kusiak & A. Marcinkowska (Eds.), *34th EARSeL Symposium. European remote sensing - new opportunities for science and practice. Abstract and Programme Book.* (pp. 317). University of Warsaw,: Warsaw.

Fun, T. T. (2005). *Integrating GIS into the geography curriculum of Hong Kong schools.* University of Hong Kong, Hong Kong.

Funke, J., Siegmund, A., Wüstenberg, S., & Viehrig, K. (2011). HEIGIS Kurzbeschreibung. Retrieved 2013-05-21, from http://kompetenzmodelle.dipf.de/de/projekte/heigis

Gandhi Kingdon, G. (2006). Teacher characteristics and student performance in India: A pupil fixed effects approach. Retrieved 2009-06-15, from www.gprg.org/pubs/workingpapers/pdfs/gprg-wps-059.pdf

García-Granero, M. (2008). Gender as Covariate in ANCOVA, Schools/Teachers as Dummy Variables, ANCOVA v. Regression? Retrieved 2012-01-10, from http://spssx-discussion.1045642.n5.nabble.com/Gender-as-Covariate-in-ANCOVA-Schools-Teachers-as-Dummy-Variables-ANCOVA-v-Regression-td1085437.html; post on October 30, 2008, 7:56 pm

Gardner, H. (2005). *Multiple Lenses on The Mind*. presented at the ExpoGestion Conference, May 25, 2005.

Garson, D. G. (2011). Factor Analysis. Retrieved 2011-05-06, from http://faculty.chass.ncsu.edu/garson/PA765/factor.htm

Garver, M. S., Williams, Z., & Taylor, S. G. (2008). Sample. Retrieved 2010-10-14, from http://findarticles.com/p/articles/mi_qa3705/is_200807/ai_n31425593/pg_5/?tag=content;col1

Gaumnitz, E. (1993). Der Amboseli-Nationalpark in Kenia: Naturschutz in der Dritten Welt. *Geographie heute*(115), 14-15.

Geißler, R., & Weber-Menges, S. (2008). Migrantenkinder im Bildungssystem: doppelt benachteiligt. *Politik und Zeitgeschichte, 49*, 14-22.

Genevois, S., & Joliveau, T. (2009). Using a Geoinformation-Based Learning Environment in Geography Education - the GeoWebExplorer Platform. *Learning with Geoinformation IV - Lernen mit Geoinformation IV*, 113-120.

Gersmehl, P. J., & Gersmehl, C. A. (2006). Wanted: A Concise List of Neurologically Defensible and Assessable Spatial Thinking Skills. *Research in Geographic Education, 8*, 5-38.

Gersmehl, P. J., & Gersmehl, C. A. (2007). Spatial Thinking by Young Children: Neurologic Evidence for Early Development and "Educability". *Journal of Geography, 106*, 181-191.

GEW. (2004). Von der Ankunft in den Schulen! Ergebnisse der GEW-Befragung zur Bildungsplanreform im Juli 2004. Retrieved 2009-12-14, from http://www.gew-bw.de/Binaries/Binary6315/von-der-ankunft.pdf?SID=899c0cc68346c690145b461abe9f7384

Gewin, V. (2004). Mapping opportunities. *Nature, 427*(22. January), 376-377.

GI. (2008). Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I. *LOG IN, 28*(150/151), Beilage.

Gladwell, M. (2008). *Outliers. The story of success*. New York: Little, Brown and Company.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Retrieved 2010-03-01, from https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence=1

Glorius, B. (2012). Internationalisierung und Heterogenisierung der Bevölkerung – Eine Herausforderung für die Gestaltung kommunaler Bildungslandschaften. In S. Maretzke (Ed.), *DGD-Jahrestagung 2011. Schrumpfend, alternd, bunter? Antworten auf den demographischen Wandel* (pp. 65-71). Bonn: DGD.

Glover, B. (2006). Demand for GIS technicians likely to soar in coming decade. Retrieved 2008-11-29, from http://www.com.edu/newsdesk/news.cfm?newsid=493

Golledge, R. G., Marsh, M. J., & Battersby, S. E. (2008). A conceptual framework for facilitating geospatial thinking. *Annals of the Association of American Geographers, 98*(2), 285-308.

Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments. Volume 3* (pp. 125-156): IEA-ETS Research Institute.

Göroğlu, R. (2013). PISA-Studie 2012. Auch Schüler mit Migrationshintergrund holen auf. Retrieved 2014-06-24, from http://mediendienst-integration.de/artikel/schueler-mit-migrationshintergrund-holen-auf.html

GPJE. (2004). *Anforderungen an nationale Bildungsstandards für den Fachunterricht in der politischen Bildung an Schulen. Ein Entwurf*. Schwalbach: Wochenschau.

Graham, J. (2007). HLM. Retrieved 2013-05-11, from myweb.facstaff.wwu.edu/~graham7/psy515/HLMnotes.ppt

Greiff, S., & Funke, J. (2008). Measuring Complex Problem Solving: The MicroDYN approach. Retrieved 2010-08-12, from http://www.psychologie.uni-heidelberg.de/ae/allg/mitarb/sg/artikel_reykjavik.pdf

Grimshaw, J., Campbell, M., Eccles, M., & Steen, N. (2000). Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practics - an international journal*(17), S11-S18.

Grube, C., Möller, A., & Mayer, J. (2007). *Dimensionen eines Kompetenzstrukturmodells zum Experimentieren*. presented at the Tagung der Fachsektion Didaktik der Biologie. Essen. Retrieved from http://www.biodidaktik.de/abstracts/vortraegeEssen/Grube+Moeller+Mayer.pdf

Gryl, I., & Kanwischer, D. (2012). Von der Kompetenz zur Performanz – Bestehende Modelle zur Kartenarbeit und Alternativen. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 154-162). Braunschweig: Westermann.

Gudovitch, Y., & Orion, N. (2001). The Carbon Cycle and Earth Systems - Studying the Carbon Cycle in Multidisciplinary Environmental Context *Science and Technology Education: Preparing Future Citizens. Proceedings of the IOSTE Symposium in Southern Europe, Cyprus, April 29 - May 2, 2001*.

Hagevik, R. A. (2003). *The effects of online science instruction using geographic information systems to foster inquiry learning of teachers and middle school science students.* Dissertation, North Carolina State University.

Haggett, P. (2004). *Geographie eine globale Synthese*. Stuttgart: Ulmer.

Hall, J. K. (2000). *Field Dependence-Independence and Computer-based Instruction in Geography.* Dissertation, Virginia Polytechnic Institute, Blacksburg.

Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education, 41*(1), 66-72.

Hannafin, R. D., Truxaw, M. P., Vermillon, J. R., & Liu, Y. (2008). Effects of Spatial Ability and Instructional Program on Geometry Achievement. *The Journal of Educational Research, 101*(3), 149-156.

Hartig, J. (2005). Analysis of hierarchical data. Retrieved 2013-05-11, from user.uni-frankfurt.de/~johartig/hlm/HLM_EARLI.pdf

Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 127-143). Heidelberg: Springer.

Haslam, A. S., & McGarty, G. (2003). *Research methods and statistics in psychology*. London: Sage.

Hattie, J. (2003). *Teachers make a difference: What is research evidence?* presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.

Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Haubrich, H. (2001). Methodenkompetenz. *Geographie heute, 19*, 2-7.

Haubrich, H. (Ed.). (1982). *Geographische Erziehung im internationalen Blickfeld*. Braunschweig: Westermann.

Haubrich, H. (Ed.). (2006). *Geographie unterrichten lernen, Die neue Didaktik der Geographie konkret* (2 ed.). München: Oldenbourg.

Haubrich, H., Reinfried, S., & Schleicher, Y. (2007). Lucerne Declaration on Geographical Education for Sustainable Development. In S. Reinfried, Y. Schleicher & A. Rempfler (Eds.), *Geographical Views on Education for Sustainable Development. Proceedings of the Lucerne-Symposium, Switzerland, July 29-31, 2007* (pp. 243-250): Geographiedidaktische Forschungen, Volume 42.

Heiken, A., & Peyke, G. (2005). GIS-Kompetenz erwerben, Das Beispiel SchulGIS. *Geographie heute, 26*(233), 30-32.

Heinze, A., Reiss, K., & Rudolph, F. (2005). Mathematics achievement and interest in mathematics from a differential perspective. *ZDM, 37*(3), 212-220.

Helmke, A. (2003). *Unterrichtsqualität. Erfassen - Bewerten - Verbessern*. Seelze: Kallmeyer.

Helmke, A., & Hosenfeld, I. (n.d.). Vera 2004. Länderkurzbericht. Retrieved 2010-04-15, from http://www.uni-landau.de/vera/downloads/Laenderkurzbericht.pdf

Hemmer, I. (1995). Geographie - kein Fach für Mädchen? Ein Überblick über den Stand der Diskussion. *Geographie und ihre Didaktik, 23*(4), 211-225.

Hemmer, I., Bayrhuber, H., Häußler, P., Hemmer, M., Hlawatsch, S., Hoffmann, L., et al. (2007). Students' interest in geoscience topics, contexts and methods. *Geographie und ihre Didaktik*(4), 185-197.

Hemmer, I., & Hemmer, M. (1997). Arbeitsweisen im Erdkundeunterricht - Ergebnisse einer empirischen Untersuchung zum Schülerinteresse und zur Einsatzhäufigkeit. In F. Frank, V. Kaminske & G. Obermaier (Eds.), *Die Geographiedidaktik ist tot, es lebe die Geographiedidaktik* (pp. 67-78). München.

Hemmer, I., & Hemmer, M. (1998). Wie beurteilen Schüler und Schülerinnen das Unterrichtsfach Geographie? – Ergebnisse einer empirischen Studie. *Geographie und Schule*(112), 40-43.

Hemmer, I., & Hemmer, M. (2002a). Mit Interesse lernen. Schülerinteresse und Geographieunterricht. *Geographie heute*(202), 2-7.

Hemmer, I., & Hemmer, M. (2002b). Wie kann ich die Interessen meiner Schüler ermitterln? *Geographie heute*(202), 10-12.

Hemmer, I., & Hemmer, M. (2007). Nationale Bildungsstandards im Fach Geographie. Genese, Standortbestimmung, Ausblick. *Geographie heute*(255/256), 2-9.

Hemmer, I., & Hemmer, M. (2013). Kompetenzmodelle. In D. Böhn & G. Obermaier (Eds.), *Wörterbuch der Geographiedidaktik. Begriffe von A – Z* (pp. 156-157). Braunschweig: Westermann.

Hemmer, I., Hemmer, M., Hüttermann, A., & Ullrich, M. (2010). Kartenauswertekompetenz – Theoretische Grundlagen und Entwurf eines Kompetenzstrukturmodells. *Geographie und ihre Didaktik*(3), 158-171.

Hemmer, I., Hemmer, M., Hüttermann, A., & Ullrich, M. (2012). Über welche grundlegenden Fähigkeiten müssen Schülerinnen und Schüler verfügen, um eine Karte auswerten zu können? Auf dem Weg zu einem Kompetenzmodell zur Kartenauswertekompetenz. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 144-153). Braunschweig: Westermann.

Hemmer, I., Hemmer, M., Kruschel, K., Neidhardt, E., Obermaier, G., & Uphues, R. (2010). Einflussfaktoren auf die kartengestützte Orientierungskompetenz von Kindern in Realräumen – Anlage eines Forschungsprojektes. *Geographie und Ihre Didaktik, 38*(2), 65-76.

Hemmer, I., Hemmer, M., Kruschel, K., Neidhardt, E., Obermaier, G., & Uphues, R. (2012). Zur Relevanz ausgewählter personenbezogener Einflussfaktoren auf die kartengestützte Orientierungskompetenz. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 64-73). Braunschweig: Westermann.

Hemmer, I., Hemmer, M., Obermaier, G., & Uphues, R. (2008). Räumliche Orientierung. Eine empirische Untersuchung zur Relevanz des Kompetenzbereichs aus der Perspektive der Gesellschaft und der Experten. *Geographie und ihre Didaktik*(1), 17-32.

Hemmer, M. (2000). *Westen ja bitte - Osten nein danke! – Empirische Untersuchungen zum geographischen Interesse von Schülerinnen und Schülern an den USA und der GUS*. Nürnberg: Geographiedidiaktische Forschungen.

Hemmer, M. (2008a). Der Kompetenzbereich "Erkenntnisgewinnung/ Methoden". Struktur und Implementierung. *Praxis Geographie*(7-8), 4-9.

Hemmer, M. (2008b). Ringvorlesung Geographie: Interesse von Schülern an einzelnen Themen, Regionen und Arbeitsweisen im Geographieunterricht, 2008-06-17. Retrieved 2010-04-15, from http://www.uni-muenster.de/Geographiedidaktik/mitarbeiter/ringvorlesungss08.html

Herodot. (2008). GIS in secondary school education: A benchmark statement. Retrieved from http://www.herodot.net/geography-benchmark.html

Herodot. (2009). Benchmark on GIS in school geography and teacher education. Retrieved 2009-07-16, from http://www.herodot.net/geography-benchmark.html

Herzig, B., & Grafe, S. (2006). *Digitale Medien in der Schule. Standortbestimmung und Handlungsempfehlungen für die Zukunft. Zusammenfassung einer Studie zur Nutzung digitaler Medien in allgemein bildenenden Schulen in Deutschland*: Telekom.

Herzig, R. (2007). GIS in der Schule - Auf dem Weg zu einer GIS-Didaktik. *Kartografische Nachrichten*(4), 199-206.

Heyden, K. (2004). GIS-Projekt: Bau einer Umgehungsstraße. *Praxis Geographie*(2), 32-34.

Hildebrandt, K. (2006). *Die Wirkung systemischer Darstellungsformen und multiperspektivischer Wissensrepräsentationen auf das Verständnis des globalen Kohlenstoffkreislaufs.* Dissertation, Christian-Albrechts-Universität, Kiel.

Hipkins, R. (2006). *The Nature of Key Competencies. A Background Paper*. Wellington: New Zealand Council for Educational Research.

Hlawatsch, S., Hildebrandt, K., Bayrhuber, H., Hansen, K.-H., & Thiele, M. (2005). *Modul 1: System Erde - Die Grundlagen. Begleittext für Lehrkräfte*. Kiel: IPN.

Hlawatsch, S., Lücken, M., Hansen, K.-H., Fischer, M., & Bayrhuber, H. (2005). Forschungsdialog: System Erde – Schlussbericht – Retrieved 2011-11-02, from http://systemerde.ipn.uni-kiel.de/Schlussbericht20_12_05-EF.pdf

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of applied quantitative methods, 3*(1), 76-83.

Hoenig, K., & Niedenzu, A. (2004). *Diercke GIS Tutor*. Braunschweig: Westermann.

Hoffmann, R. (2000). Leitbilder für den Geographieunterricht: Bestand - Bedarf? *Geographie und ihre Didaktik, 28*(3), 120-130.

Hofmann, C. (2013). Die weisse Massai. Retrieved 2013-08-12, from http://www.massai.ch

Höhnle, S., Schubert, J. C., & Uphues, R. (2010). The frequency of GI(S) use in the geography classroom – Results of an empirical study in German secondary schools. In T. Jekel, A. Koller, K. Donert & R. Vogler (Eds.), *Learning with Geoinformation V - Lernen mit Geoinformation V* (pp. 148-158). Berlin: Wichmann.

Höhnle, S., Schubert, J. C., & Uphues, R. (2011). Barriers to GI(S) use in schools – A comparison of international empirical results. In T. Jekel, A. Koller, K. Donert & R. Vogler (Eds.), *Learning with GI 2011* (pp. 124-133): Wichmann.

Hooey, C. A., & Bailey, T. J. (2005). Journal Writing and the Development of Spatial Thinking Skills. *Journal of Geography, 104*(6), 257-261.

Horx Zukunftsinstitut GmbH. (2011). Delphi-Methode. Retrieved 2014-06-27, from http://www.horx.com/zukunftsforschung/Docs/02-M-09-Delphi-Methode.pdf

Hox, J. (1998). Multilevel Modeling: When and Why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147-154). New York: Springer.

Huang, W.-M. (n.d.). ANCOVA (Analysis of Covariance). Retrieved 2011-05-10, from http://www.lehigh.edu/~wh02/ancova.html

Hüttermann, A. (2004). *Kartographische Kompetenzen im Geographieunterricht allgemein bildender Schulen*. presented at the Vortrag auf der Intergeo 2004 Stuttgart. Retrieved from http://www.intergeo.de/archiv/2004/Huettermann.pdf

Hüttermann, A. (2005). Kartenkompetenz: Was sollen Schüler können?. *Praxis Geographie, 35*(11), 4-8.

Hüttermann, A. (2012). Von der „Einführung in das Kartenverständnis" zu „Kartenkompetenz": Der schillernde Begriff der Kartendidaktik. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 22-32). Braunschweig: Westermann.

Hüttermann, A., Fichtner, U., & Herzig, R. (2010). Regelkreis zur Gewinnung kartographischer Kompetenz. *Geographie und Ihre Didaktik, 38*(2), 77-88.

Huynh, N. T. (2008). Measuring and Developing Spatial Thinking Skills in Students. In T. Jekel, A. Koller & K. Donert (Eds.), *Learning with Geoinformation III - Lernen mit Geoinformation III* (pp. 116-125). Heidelberg: Wichmann.

Huynh, N. T. (2009). *The role of geospatial thinking and geographic skills in effective problem solving with GIS: K-16 education.* Wilfrid Laurier University, Waterloo, Ontario.

IBO. (2005). Diploma programme geography. First examinations 2005. Retrieved 2010-03-29, from http://www.scotch.wa.edu.au/upload/pages/ib-diploma-curriculum/geography-guide.pdf

Ibrahim, F. (2003). Desertifikation und Nachhaltigkeit in Ostafrika. Das Beispiel der Maasai Nord-Tansanias. *Geographische Rundschau, 55*(7-8), 4-9.

Institut Beatenberg. (n.d.). Kompetenzraster. Retrieved 2014-06-23, from http://www.institutbeatenberg.ch/images/pdf/kompetenzraster/kompetenzraster.pdf

Institut für Physik und ihre Didaktik. (2013). Design Based Research. Retrieved 2014-06-27, from http://www.physikdidaktik.uni-koeln.de/11287.html

Instituto Cervantes. (2014). Sample examination papers and audio materials. Retrieved 2014-06-23, from http://diplomas.cervantes.es/en/information/sample_examination_papers_audios.html

Ipara, H. (2002). Towards cultural tourism development around the Kakamega Forest Reserve, Kenya. In J. S. Akama & P. Sterry (Eds.), *Cultural tourism in Africa: strategies for the new millennium, Proceedings of the ATLAS Africa International Conference, December 2000, Mombasa, Kenya* (pp. 95-107). Arnhem: ATLAS.

Jahn, M., & Siegmund, A. (2012). Der potentielle Beitrag von Fernerkundungsdaten zur Bildung für nachhaltige Entwicklung. In I. Hemmer, M. Müller & M. Trappe (Eds.), *Tagung „Rio +20 Nachhaltigkeit neu denken?". Abstractband* (pp. 40). Eichstätt: Katholische Universität Eichstätt-Ingolstadt.

jba/AP/AFP. (2007). Kenia-Wahl: Unruhen und Plünderungen. Retrieved 2013-08-12, from http://www.focus.de/politik/ausland/kenia-wahl_aid_231065.html

Jo, I., & Bednarz, S. W. (2008). *Bringing Spatial Thinking into My Classroom*. presented at the 2008 NCGE Annual Meeting, Dearborn.

Jo, I., & Bednarz, S. W. (n.d.). A taxonomy of spatial thinking. Retrieved 2011-01-13, from http://www.teachspatial.org/taxonomy-spatial-thinking

Joachim, J. (2006). WebGIS in der Schule. *Geographie aktuell*(5), 26-30.

Joachim, J. (2007). Leitlinie zum Einsatz von WebGIS im Unterricht. Retrieved 2014-05-04, from http://www.lehrer-online.de/webgis.php

Job, H., & Metzler, D. (2003). Tourismusentwicklung und Tourismuspolitik in Ostafrika. *Geographische Rundschau, 55*(7-8), 10-17.

Johansson, T. (2003). *GIS in Teacher Education – Facilitating GIS Applications in Secondary School Geography.* Paper presented at the ScanGIS 2003 - The 9th Scandinavian Research Conference on Geographical Information Science, June 4-6 2003, Espoo, Finland.

Johansson, T. (Ed.). (2006). *Geographical Information Systems Applications for Schools – GI-SAS*: University of Helsinki.

Johansson, T., & Pellikka, P. (2005). *Interactive Geographical Information Systems (GIS) Applications for European Upper Secondary Schools*. presented at the Recent Research Developments in Learning Technologies, III International Conference on Multimedia & ICT's in Education, FORMATEX.

Kainz, D., & Ossimitz, G. (2002). *Can students learn stock-flow-thinking? An empirical investigation*. presented at the System Dynamics Conference, Palermo.

Kali, Y. (2003). A virtual journey within the rock-cycle: a software kit for the development of systems-thinking in the context of the earth's crust. *Journal of Geoscience Education, 51*(2), 165-170.

Kali, Y., & Orion, N. (1996). Spatial Abilities of High-School Students in the Perception of Geologic Structures. *Journal of Research in Science Teaching, 33*(4), 369-391.

Kali, Y., Orion, N., & Eylon, B.-S. (2003). Effect of Knowledge Integration Activities on Students' Perception of the Earth's Crust as a Cyclic System. *Journal of Research in Science Teaching, 40*(6), 545-565.

Kaminske, V. (1996). Vernetztes Denken im Unterricht - Der geosystemare Ansatz in einer Klasse 11. *Geographie und Schule*(102), 36-43.

Kao, G., & Thompson, J. S. (2003). Racial and Ethnic Stratification in Educational Achievement and Attainment. *Annual Review of Sociology, 23*, 417-442.

Kapulnik, E., Orion, N., & Ganiel, U. (2004). *In-Service Training programs in science & technology (IT) diffusion: K-12 student and educator conceptualizations*. presented at the National Association for Research in Science Teaching, (NARST) Symposium.

Kasperidus, H. D., Langfelder, H., & Biber, P. (2006). *Comparing Systems Thinking Inventory Task Performance in German Classrooms at High School and University Level*. presented at the International Conference of the System Dynamics Society.

Kaufhold, M. (2006). *Kompetenz und Kompetenzerfassung. Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Keiper, T. A. (1999). GIS for Elementary Students: An Inquiry Into a New Approach to Learning Geography. *Journal of Geography*(98), 47-59.

Keranen, K., & Kolvoord, R. (2008). *Making Spatial Decisions Using GIS. Level 4*. Redlands: ESRI Press.

Kerski, J. J. (2000). *The Implementation and Effectiveness of GIS Technology and Methods in Secondary Education.* Dissertation, University of Colorado.

Kerski, J. J. (2003). The implementation and effectiveness of geographic information systems technology and methods in secondary education. *Journal of Geography, 102*, 128-137.

Kerski, J. J. (2006). *Exploring Africa with the MapMachine. Lessons: Cruising along Africa's Great Arcs – The equator and the prime meridian. Eco-Regions of Africa. Great Lakes of Africa. Islands of Africa. Moving to Africa. Seasons in Africa. Snow in Africa.*: ESRI.

Kerski, J. J. (2008a). Geographic Information Systems in Education. In J. P. Wilson & S. A. Fotheringham (Eds.), *The handbook of geographic information science* (pp. 540-556). Malden: Blackwell.

Kerski, J. J. (2008b). Geography Awareness Week resources, GIS Day, Revised Elections Data and Lesson, Teaching History with GIS, Taiwan GIS. In EDGIS Mailing List (Ed.).

Kerski, J. J. (2009). Joseph Kerski. Retrieved 2009-09-13, from http://www.josephkerski.com/

Kersting, R. (2002). Wo sind die Mädchen? Erste Ergebnisse einer Befragung von Schülerinnen und Schülern von Erdkundekursen in der Sek. II. *Geographie heute*(202), 20-21.

Kestler, F. (2002). *Einführung in die Didaktik des Geographieunterrichts*. Bad Heilbrunn: Klinkhardt.

Kfir, D. (1988). Achievements and aspirations among boys and girls in high school: a comparison of two Israeli ethnic groups. *American Educational Research Journal, 25*(2), 213-236.

Kinzel, M. R. (2009). *Using educational tools and integrative experiences via geovisualizations that incorporate spatial thinking, real world science and ocean literacy standards in the classroom: a case study examined.* Oregon State University, Corvallis.

Klaus, D. (1985). Allgemeine Grundlagen des systemtheoretischen Ansatzes. *Geographie und Schule*(33), 1-8.

Klein, U. (2005). Räumliche Prozesse werden anschaulich, Schüler bewerten GIS im Unterricht. *ArcAktuell*(3), 27.

Klein, U. (2007). *Geomedienkompetenz. Untersuchung zur Akzeptanz und Anwendung von Geomedien im Geographieunterricht unter besonderer Berücksichtigung moderner Informations- und Kommunikationstechniken.* Dissertation, Christian-Albrechts-Universität, Kiel.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (2007). *Zur Entwicklung nationaler Bildungsstandards – Expertise*. Bonn: Bundesministerium für Bildung und Forschung.

Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik, 47*, 179-200.

Klieme, E., & Leutner, D. (2006a). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik, 52*(6), 876-903.

Klieme, E., & Leutner, D. (2006b). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Überarbeitete Fassung des Antrags an die DFG auf Einrichtung eines Schwerpunktprogramms. Retrieved 2009-01-04, from http://kompetenzmodelle.dipf.de/images/sppfiles/files/antrag_spp_kompetenz modelle.pdf

Klieme, E., & Maichle, U. (1991). *Erprobung eines Modellbildungssystems im Unterricht*. Bonn: Institut für Test- und Begabungsforschung.

Klieme, E., & Maichle, U. (1994). *Modellbildung und Simulation im Unterricht der Sekundarstufe I*. Bonn: Institut für Bildungsforschung.

Kline, R. B. (2000). D.4 Requirements. Retrieved 2011-10-14, from http://psychology.concordia.ca/fac/kline/Supplemental/latent_d.html#D.4%20Requirements

KMK. (2003a). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss*. München: Wolters Kluwe.

KMK. (2003b). *Bildungsstandards im Fach Mathematik für den Mittleren Bildungsabschluss*. München: Wolters Kluwer.

KMK. (2004a). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. München: Wolters-Kluwer.

KMK. (2004b). *Bildungsstandards im Fach Physik für den Mittleren Bildungsabschluss*. München: Wolters Kluwer.

KMK. (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. München: Luchterhand.

KMK. (2005b). Einheitliche Prüfungsanforderungen in der Abiturprüfung Geographie. Retrieved 2010-03-22, from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Geographie.pdf

KMK. (2008). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Retrieved 2010-03-22, from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile.pdf

KMK. (2014). The Education System in the Federal Republic of Germany. Retrieved 2014-02-14, from http://www.kmk.org/information-in-english/the-education-system-in-the-federal-republic-of-germany.html

Koballa, T. (n.d.). Framework for the affective domain in science education. Retrieved 2014-03-16, from http://serc.carleton.edu/NAGTWorkshops/affective/framework.html

Köck, H. (1977). Ziele des Geograhieunterrichts seit 1945. Versuch eines Überblicks der Grunddimensionen der geographischen Zielsetzung in der Primarstufe, Sekundarstufe I und II. *Hefte zur Fachdidaktik der Geographie*(1), 3-53.

Köck, H. (1985). Systemdenken - geographiedidaktische Qualifikation und unterrichtliches Prinzip. *Geographie und Schule, 7*(33), 1-8.

Köck, H. (1989). Aufgabe und Aufbau des Geographieunterrichts. *Geographie und Schule*(57), 11-25.

Köck, H. (1993). Raumbezogene Schlüsselqualifikationen - der fachimmanente Beitrag des Geographieunterrichts zum Lebensalltag des Einzelnen und Funktionieren der Gesellschaft. *Geographie und Schule, 15*(8), 14-22.

Köck, H. (1997). Raumverhaltenskompetenz in der Kritik und die Frage nach möglichen Leitzielalternativen. In F. Franke, V. Kaminske & G. Obermaier (Eds.), *Die Geographiedidaktik ist tot, es lebe die Geographiedidaktik. Festschrift zur Emeritierung von Josef Birkenhauer.* (pp. 17-39). München.

Köck, H. (1999). Systemische Welt - Systemische Geographie, Notwendigkeit und Möglichkeit eines angemessenen Weltzugriffs. In H. Köck (Ed.), *Geographieunterricht und Gesellschaft, Vorträge des gleichnamigen Symposiums vom 12.-15. Okt. 1998 in Landau* (pp. 163-181). Nürnberg.

Köck, H. (2003). Dilemmata der (geographischen) Umwelterziehung (Teil II). *Geographie und ihre Didaktik, 31*(2), 61-79.

Köck, H. (2005a). Leitziel. In H. Köck & D. Stonjek (Eds.), *ABC der Geographiedidaktik* (pp. 161). Köln.

Köck, H. (2005b). Raumverhaltenskompetenz. In H. Köck & D. Stonjek (Eds.), *ABC der Geographiedidaktik* (pp. 210). Köln.

Köck, H. (Ed.). (1986). *Grundlagen des Geographieunterrichts*. Köln.

Köck, H., & Rempfler, A. (2004). *Erkenntnisleitende Ansätze - Schlüssel zur Profilierung des Geographieunterrichts mit erprobten Unterrichtsvorschlägen*. Köln: Aulis-Verl. Deubner.

Köck, H., & Stonjek, D. (2005). *ABC der Geographiedidaktik*. Köln: Aulis-Verl. Deubner.

Kohl, L. (2008). *Arbeitsmaterial zur Untersuchung "Kompetenzorientierter Informatikunterricht mit der visuellen Programmiersprache Puck"*. Jena: Friedrich-Schiller-Universität.

Kohlmann, T., & Moock, J. (2009). 5 How to analyze your data. In D. Stengel, M. Bhandari & B. Hanson (Eds.), *Handbook – Statistics and data management. A practical guide for orthopaedic surgeons* (pp. 93-109): AO Publishing.

Kollar, I. (2012). *Die Satellitenbild-Lesekompetenz. Empirische Überprüfung eines theoriegeleiteten Kompetenzstrukturmodells für das „Lesen" von Satellitenbildern.* Dissertation, Heidelberg University of Education, Heidelberg.

Koller, A. (2005). Web-GIS – ein Werkzeug für den Geographieunterricht, Begriffsklärung und Unterrichtsbeispiele. *Geographie heute, 26*(233), 12-17.

Kramer, B. (2005). *Mentale Integration von Text und Bild beim Lernen mit Multimedia am Beispiel der olfaktorischen Signaltransduktion.* Dissertation, Christian-Albrechts-Universität, Kiel.

Krause, T. (2004). Digitaler Kinderstadtteilplan, Ein GIS-Projekt aus dem Schulumfeld für die Orientierungsstufe. *Praxis Geographie, 34*(2), 13-15.

Krause, T. (2005). Arbeiten mit digitalen Karten - Kartenkompetenz durch die Nutzung von Geoinformationssystemen. *Praxis Geographie*(11), 30-33.

Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies.* Los Angeles: California State University.

Kreger Silverman, L. (2000). *Identifying Visual-Spatial and Auditory-Sequential Learners: A Validation Study*. presented at the Henry B. and Jocelyn Wallace National Research Symposium on Talent Development, May 19, 2000.

Kreger Silverman, L. (2005). *Upside-down brilliance: The visual spatial learner*. presented at the Queensland Association for Gifted and Talented Children, March 12, 2005.

Kriz, W. C. (2000). *Lernziel: Systemkompetenz Planspiele als Trainingsmethode*. Göttingen: Vandenhoeck & Ruprecht.

Kross, E. (1992). Von der Inwertsetzung zur Bewahrung der Erde. *Geographie heute, 13*(100), 57-62.

Kross, E. (1994). Die Erde bewahren - die neue Leitidee für den Geographieunterricht. In M. Flath & G. Fuchs (Eds.), *Die Erde bewahren - Fremdartigkeit verstehen und respektieren, Erstes Gothaer Forum zum Geographieunterricht, 26./27. November 1993 in Gotha* (pp. 16-23). Gotha: Justus Perthes Verlag.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., et al. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max Planck Institut für Bildungsforschung.

Kuntze, S., Lindmeier, A., & Reiss, K. (2008). "*Using models and representations in statistical contexts*" *as a sub-competency of statistical literacy – Results from three empirical studies*. presented at the 11th International Congress on Mathematical Education (ICME 11), University of Monterrey, Mexico.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: understanding achievement in U.S. schools. *Educational researcher, 35*(7), 3-12.

Laib, N. A. (2007). *Promoting Academic Success for Limited English Proficient Students.* Master thesis, The Evergreen State College.

Lamnek, S. (2005). *Qualitative Sozialforschung*. Weinheim: Beltz.

Langer, W. (2005). Workshop Mehrebenenanalyse: Statistische Grundlagen und Anwendungspraxis mit MLA 4 und HLM 5. Retrieved 2011-05-12, from http://www.soziologie.uni-halle.de/langer/pdf/papers/mla_workshop_linz_12052005.pdf

Learning Africa. (n.d.). Eco-tourism. Retrieved 2007-12-11, from http://www.learningafrica.org.uk/poverty_activities.htm

Lecher, T. (1997). *Die Umweltkrise im Alltagsdenken*. Weinheim: Beltz.

Lee, J. (2009). Self-constructs and anxiety across cultures. Retrieved 2010-04-15, from http://www.etsliteracy.net/Media/Research/pdf/RR-09-12.pdf

Lee, J. W. (2005). *Effect of GIS learning on spatial ability.* Dissertation, Texas A & M University.

Lee, J. W. (n.d.). Spatial Skills Test. Retrieved 2008-09-02, from http://www.aag.org/tgmg_old/sst.htm

Lee, M.-H., Johanson, R. E., & Tsai, C.-C. (2007). Exploring Taiwanese High School Students' conceptions of and approaches to learning science through a structural equation modeling analysis. *Science Education*, 191-220.

Lee, P. (1996). Cognitive development in bilingual children: a case for bilingual instruction in early childhood education. *The bilingual research journal, 20*(3 & 4), 499-522.

Lensment, L. (2002). *Systematische und problemorientierte Erarbeitung von chemischen Sachverhalten im Rahmen eines multimedialen Lernprogramms, Eine empirische Studie zum Kapitel Wasser der Internetvorlesung CHEMnet.* Dissertation, Christian-Albrechts-Universität, Kiel.

Lenz, T. (2003). Leistungsüberprüfung und Leistungsbewertung im bilingualen Geographieunterricht. *Geographie und Schule, 25*(143), 38-45.

Lenz, T. (2004). Förderung der Lese- und Methodenkompetenz. *Geographie heute*(224), 5-11.

Lenz, T. (2012). Neue Aufgabenkultur im Geographieunterricht am Beispiel der Kartenauswertekompetenz. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 192-203). Braunschweig: Westermann.

Leopold, C., Górska, R. A., & Sorby, S. A. (2001). International Experiences in Developing the Spatial Visualization Abilities of Engineering Students. *Journal for Geometry and Graphics, 5*(1), 81-91.

Lethmate, J. (2001). Ökologie und Erdkunde: Geographiedidaktische Selbstillusionierungen. *Geographie und Schule*(132), 37-43.

LeVasseur, M. L. (1999). Students' Knowledge of Geography and Geography Careers. *Journal of Geography, 98*(6), 265-271.

Levin, H. M. (1995). Cost-effectiveness Analysis. In M. Carnoy (Ed.), *International Encyclopedia of Economics of Education* (pp. 381- 386). Oxford: Pergamon.

Levin, H. M. (2001). Waiting for Godot: Cost-Effectiveness Analysis in Education. *New Directions for Evaluation*(90), 55-68.

Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research, 28*(4), 612-625.

LfS. (2012). (Ed.) Mit Kompetenzrastern dem Lernen auf der Spur. Retrieved 2013-05-29, from http://www.ls-bw.de/Handreichungen/pub_online/NL04_Mit_Kompetenzrastern_dem_Lernen_auf_der_Spur.pdf

LfS. (n.d.). Publikationen online. Retrieved 2013-05-29, from http://www.ls-bw.de/Handreichungen/pub_online/

Li, P., Arbarbanell, L., & Papafragou, A. (2005). Spatial reasoning skills in Tenejapan Mayans *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (pp. 1272-1277): Lawrence Erlbaum Associates, Inc.

Lichtenstein, M. J., Owen, S. V., Blalock, C. L., Liu, Y., Ramirez, K. A., Pruski, L. A., et al. (2008). Psychometric Reevaluation of the Scientific Attitude Inventory-Revised (SAI-II). *Journal of Research in Science Teaching, 45*(5), 600-616.

Linacre, J. M. (1994). Sample Size and Item Calibration Stability. *Rasch Measurement Transactions, 7*(4), 328. Retrieved from http://www.rasch.org/rmt/rmt74m.htm

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106. Retrieved from http://www.winsteps.com/a/linacre-optimizing-category.pdf

Lindau, A.-K. (2010). Kartieren – eine Geländemethode im Geographieunterricht. *Geographie und Ihre Didaktik, 38*(2), 109-117.

Lindmeier, A., Kuntze, S., & Reiss, K. (2007). *Representations of data and manipulations through reduction – competencies of German secondary students*. presented at the IASE/ISI Satellite Conference on Statistical Education. Guimarães, Portugal.

Link, C. (Writer). (2001). Nirgendwo in Afrika, *based on the novel by Stefanie Zweig*. München: Constantin Film.

Liu, X. (2004). Using Concept Mapping for Assessing and Promoting Relational Conceptual Change in Science. *Science Education, 88*(3), 373-396.

Lloret Ivorra, E. M., Ribas, R., & Wiener, B. (2009). *Con Gusto. Lehr- und Arbeitsbuch mit 2 Audio-CDs. A1*: Klett.

Loi, M., & Ronsivalle, B. G. (2005). *A Particular Aspect of Cost Analysis in Distance Education: Time*. presented at the European Distance and E-Learning Network Conference, June 20-23 2005, Helsinki.

Lücken, M., Hlawatsch, S., & Raack, N. (2005). Kompetenzentwicklung im fachübergreifenden Unterricht im Kontext der Geowissenschaften: die Entwicklung einer Methode zur Erhebung von Systemkompetenz. Retrieved 2010-03-25, from http://www.biodidaktik.de/abstracts/vortraege/21.pdf

Ludford, P. J. (2003). ANCOVA (Analysis of Covariance). Retrieved 2011-05-10, from http://www-users.cs.umn.edu/~ludford/Stat_Guide/ANCOVA.htm

Lund, J. J., & Sinton, D. S. (2007). Critical and creative visual thinking. In D. S. Sinton & J. J. Lund (Eds.), *Understanding Place. GIS and Mapping Across the Curriculum* (pp. 1-17). Redlands: ESRI Press.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

Magidson, J., & Vermunt, J. K. (n.d.). Latent Class Models. Retrieved 2011-10-12, from http://www.statisticalinnovations.com/articles/lcmodels.pdf

Maier, W. (1998). *Grundkurs Medienpädagogik und Mediendidaktik. Ein Studien- und Arbeitsbuch*. Weinheim: Beltz.

Maierhofer, M. (2001). *Förderung des systemischen Denkens durch computergestützten Biologieunterricht.* Dissertation, Universität Salzburg, Salzburg.

Makonjio Okello, M., & Warui Kiringe, J. (2004). Threats to Biodiversity and their Implications in Protected and Adjacent Dispersal Areas of Kenya. *Journal of Sustainable Tourism, 12*(1), 55-69.

Malone, L. (2003). An Example of using GIS to think spatially. In National Research Council (Ed.), *Learning to think spatially. GIS as a support system in the K-12 curriculum* (pp. 237-240). Washington D.C.: National Academies Press.

Malone, L., Palmer, A. M., & Voigt, C. L. (2002). *Mapping our World. GIS lessons for educators*. Redlands.

Malone, L., & Wilson, J. (n.d.). *Memo From the White House: Investigating Africa. Teacher's Guide*.

Mann, R. L. (2005). *The identification of gifted students with spatial strengths: an exploratory study.* University of Connecticut.

Marcello, J. S. (2009). A Proposal for Assessment in Geography Education. *Journal of Geography, 108*(4-5), 226-232.

Mark, D. M., & Turk, A. G. (2003). Landscape Categories in Yindjibarndi: Ontology, Environment, and Language. *COSIT 2003, Lecture Notes in Computer Science No. 2825*, 31-49.

Marlow, K. R. (2008). *Closing the Achievement Gap for English Language Learners: A Comparison of Language Arts/ESOL and One-Way Developmental Bilingual Programs.* Dissertation, College of Education, University of Central Florida, Orlando.

Marsh, M. J., Golledge, R. G., & Battersby, S. E. (2007). Geospatial Concept Understanding and Recognition in G6-College Students: A Preliminary Argument for Minimal GIS. *Annals of the Association of American Geographers, 97*(4), 696-712.

Marshall, C., & Rossman, G. B. (1999). *Designing Qualitative Research* (3rd ed.). Thousands Oak: Sage.

Marshall, V. (2004). Assessing the possible local community benefits from ecotourism operations in Kenya. Retrieved 2008-02-03, from http://www.uga.edu/juro/2004/marshall.pdf

Martin, M. O., Mullis, I. V. S., Foy, P., with:, i. c., Olson, J. F., Erberber, E., et al. (2008). *TIMSS 2007 International Science Report*: IEA.

Maths-Statistics-Tutor. (2010). Performing Normality in PASW (SPSS). Retrieved 2013-08-23, from http://www.maths-statistics-tutor.com/normality_test_pasw_spss.php

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung* (pp. 177-186). Heidelberg: Springer.

Medienpädagogischer Forschungsverbund Südwest. (2008). *JIM-Studie 2008: Jugend, Information, (Multi-)Media. Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Stuttgart.

Medienpädagogischer Forschungsverbund Südwest. (2012). *JIM 2012: Jugend, Information, (Multi-)Media. Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Stuttgart.

MEI. (2007). Spearman's rank correlation. Retrieved 2011-08-05, from http://www.mei.org.uk/files/pdf/Spearmanrcc.pdf

mer/dpa. (2007). Deutsche Bildungspolitik: Uno-Schulinspektor übt harsche Kritik. Retrieved 2010-03-22, from http://www.spiegel.de/schulspiegel/wissen/0,1518,468644,00.html

Meyer, C. (2003). *Bedeutung, Wahrnehmung und Bewertung des bilingualen Geographieunterrichts. Studien zum zweisprachigen Erdkundeunterricht (Englisch) in Rheinland-Pfalz.* Dissertation, Universität Trier.

Meyer, C. (2006). Methodenpass und Metakognition. In H. Haubrich (Ed.), *Geographie unterrichten lernen. Die neue Didaktik der Geographie konkret* (pp. 232-233). München.

Meyer, J. W., Butterick, J., Olkin, M., & Zack, G. (1999). GIS in the K-12 curriculum: a cautionary note. *Professional Geographer, 51*(4), 571-578.

Meyer, T. (1995). Tourismus in Kenia. *Geographie heute, 16*(136), 16-19.

mik/AP/Reuters/dpa. (2007). Kenia-Wahl: Die Stimmung kippt - Unruhen im ganzen Land. Retrieved 2013-08-12, from http://www.spiegel.de/politik/ausland/kenia-wahl-die-stimmung-kippt-unruhen-im-ganzen-land-a-525853.html

Mililani Mauka Elementary School. (n.d.). Affective Education and P.E. Retrieved 2014-03-16, from http://www.milmauka.k12.hi.us/documents/physed/affect.htm

Milson, A. J., Demirci, A., & Kerski, J. J. (Eds.). (2012). *International Perspectives on Teaching and Learning with GIS in Secondary Schools*: Springer.

Milson, A. J., & Earle, B. D. (2007). Internet-based GIS in an inductive learning environment: a case study of ninth-grade geography students. *Journal of Geography*(106), 227-237.

Ministry of Education Youth and Culture Jamaica. (2004). GIS in schools. Retrieved 2007-06-05, from http://www.moec.gov.jm/projects/gissep/about.htm

Ministry of Tourism and Wildlife. (2007a). Environmental impact assessment (EIA) guidelines for the tourism sector in Kenya. Retrieved 2014-05-04, from http://www.dlist-asclme.org/sites/default/files/doclib/EIA%20Guidlines%20kenya.pdf

Ministry of Tourism and Wildlife. (2007b). Facts and Figures. Retrieved 2008-01-04, from http://www.tourism.go.ke/ministry.nsf/doc/Facts%20&%20figures%202007.pdf/$file/Facts%20&%20figures%202007.pdf

MKBB. (2002). Rahmenlehrplan Geografie Sekundarstufe I. Retrieved 2007-10-10, from http://www.bildung-brandenburg.de/fileadmin/bbs/unterricht_und_pruefungen/rahmenlehrplaene/sekundarstufe_I/rahmenlehrplaene/S1-Geografie.pdf

MKBL. (2006). Rahmenlehrplan für die Sekundarstufe I. Jahrgangsstufe 7-10. Hauptschule, Realschule, Gesamtschule, Gymnasium - Geografie. Retrieved 2007-10-09, from http://www.berlin.de/imperia/md/content/sen-bildung/schulorganisation/lehrplaene/sek1_geografie.pdf

MKBR. (2006a). Welt-Umweltkunde, Geschichte, Geografie, Politik Bildungsplan für das Gymnasium Jahrgangsstufe 5 - 10. Retrieved 2007-10-10, from http://lehrplan.bremen.de/sek1/wuk/wuk_gy_5bis10/download

MKBR. (2006b). Welt-Umweltkunde, Geschichte, Geografie, Politik Bildungsplan für die Sekundarschule Jahrgangsstufe 5 - 10. Retrieved 2007-10-10, from http://lehrplan.bremen.de/sek1/wuk/wuk_sek_5bis10/download

MKBW. (2004a). Bildungsplan Allgemeinbildendes Gymnasium. Retrieved 2007-07-10, from www.bildung-staerkt-menschen.de/service/downloads/Bildungsplaene/Gymnasium/Gymnasium_Bildungsplan_Gesamt.pdf

MKBW. (2012). Die Gemeinschaftsschule in Baden-Württemberg. Retrieved 2013-05-29, from http://www.geb-bretten.de/20121018_Information.pdf

MKBW (Ed.). (2004b). *Bildungsplan Realschule*. Stuttgart.

MKBW, & LfS. (n.d.). Bildungsplanreform 2015/2016. Retrieved 2013-05-29, from http://www.arge-tuebingen.de/downloads/ARGE%20Tue%20ohne%20Animation_KM_130304_Bildungsplanreform_2015.pdf

MKBY. (2001). Lehrplan Realschule R6. Retrieved 2007-10-10, from www.isb.bayern.de/isb/index.asp?MNav=0&QNav=4&TNav=0&INav=0&Fach=&LpSta=6&STyp=5

MKBY. (2004). Lehrplan für das Gymnasium in Bayern (G8). Retrieved 2007-10-10, from http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26546, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26335, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26303, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26283, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26481, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26547, http://www.isb-gym8-lehrplan.de/contentserv/3.1/g8.de/index.php?StoryID=26548

MKHE. (n.d.-a). Lehrplan Erdkunde Bildungsgang Realschule. Retrieved 2007-10-17, from http://lernarchiv.bildung.hessen.de/reposit2/12311/LPRealErdkunde.pdf?user_id=Anonymous+User&is_viewable=1&digest=2b8764425b8a2aeabb74c40018f35b51

MKHE. (n.d.-b). Lehrplan Erdkunde Gymnasialer Bildungsgang Jahrgangsstufen 5 bis 13. Retrieved 2007-10-17, from http://www.kultusministerium.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/HKM_Internet/med/741/74150e9f-ba45-b901-be59-2697ccf4e69f,22222222-2222-2222-2222-222222222222,true.pdf

MKHH. (2003). Rahmenplan Geographie BILDUNGSPLAN HAUPTSCHULE UND REALSCHULE SEKUNDARSTUFE I. Retrieved 2007-10-10, from http://lbs.hh.schule.de/bildungsplaene/Sek-I_HR/GEO_HR_SEKI.PDF

MKHH. (2004). Rahmenplan Geographie BILDUNGSPLAN ACHTSTUFIGES GYMNASIUM SEKUNDARSTUFE I. Retrieved 2007-10-10, from http://lbs.hh.schule.de/bildungsplaene/Sek-I_Gy8/GEO_Gy8.pdf

MKMV. (2002a). Rahmenlehrplan Geographie - Gymnasium, Integrierte Gesamtschule 7-10. Retrieved 2007-10-09, from http://www.bildungsserver-mv.de/download/rahmenplaene/rp-geografie-7-10-gym-02.pdf

MKMV. (2002b). Rahmenplan Geographie - Regionale Schule, Verbundene Haupt- und Realschule, Hauptschule, Realschule, Integrierte Gesamtschule. Retrieved 2007-10-09, from http://www.bildungsserver-mv.de/download/rahmenplaene/rp-geografie-7-10-reg.pdf

MKNS. (2008a). *Kerncurriculum für das Gymnasium Schuljahrgänge 5 - 10 Erdkunde*. Hannover.

MKNS. (2008b). *Kerncurriculum für die Realschule Schuljahrgänge 5-10 Erdkunde*. Hannover: Niedersächsischer Bildungsserver (NIBIS).

MKNW (Ed.). (1993). *Richtlinien und Lehrpläne: Erdkunde. Realschule*. Düsseldorf: Verlagsgesellschaft Ritterbach.

MKNW (Ed.). (2004). *Sekundarstufe I Gymnasium. Richtlinien und Lehrpläne. Erdkunde*. Düsseldorf: Ritterbach Verlag.

MKRP. (n.d.). Lehrplan Erdkunde (Klassen 7-9/10). Hauptschule, Realschule, Gymnasium, Regionale Schule. Retrieved 2010-04-19, from http://lehrplaene.bildung-rp.de/lehrplaene-nach-faechern.html?tx_abdownloads_pi1%5Baction%5D=getviewdetailsfordownload&tx_abdownloads_pi1%5Buid%5D=200&tx_abdownloads_pi1%5Bcategory_uid%5D=89&tx_abdownloads_pi1%5Bcid%5D=5786&cHash=bdc1f1311fe4730d37ca6e709e304812

MKSA. (1999). *Rahmenrichtlinien Sekundarschule Schuljahrgänge 7-10 Geographie*. Magdeburg.

MKSA. (2003). *Rahmenrichtlinien Gymnasium Geographie Schuljahrgänge 5-12*. Magdeburg.

MKSH. (1997). Lehrplan für die Sekundarstufe I der weiterführenden allgemeinbildenden Schulen Hauptschule, Realschule, Gymnasium. Erdkunde. Retrieved 2007-10-15, from http://lehrplan.lernnetz.de/intranet1/links/materials/1107161157.pdf

MKSL. (2002). *Achtjähriges Gymnasium Lehrplan Erdkunde Klassenstufe 7*.

MKSL. (n.d.). Erweiterte Realschule 7. Bildungsgang Mittlerer Bildungsabschluss. Erdkunde. Retrieved 2010-04-19, from http://www.saarland.de/dokumente/thema_bildung/ERSLp-07.pdf

MKSN (Ed.). (2004a). *Lehrplan Gymnasium Geographie*. Dresden.

MKSN (Ed.). (2004b). *Lehrplan Mittelschule Geographie*. Dresden.

MKTH. (1999a). *Lehrplan für das Gymnasium. Geographie*. Erfurt.

MKTH. (1999b). *Lehrplan für die Regelschule und für die Förderschule mit dem Bildungsgang der Regelschule. Geographie*. Erfurt.

Montelllo, D. O., Lovelace, K. L., Golledge, R., & Self, C. M. (1999). Sex-related differences and similarities in Geographic and Environmental Spatial abilities. *Annals of the Association of American Geographers, 89*(3), 515-534.

Moosbrugger, H., & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.

Mosimann, T. (2007). Einführung in die Landschaftsökologie: der ökologische Blick auf die Landschaft. In H. Gebhardt, R. Glaser, U. Radtke & P. Reuber (Eds.), *Geographie physische Geographie und Humangeographie* (pp. 486-495). Heidelberg: Spektrum.

Moysés, S. J. (2000). *Oral health and healthy cities: an analysis of intra-urban differentials in oral health outcomes in realtion to "healthy cities" policies in Curitiba, Brazil.* University College London, London, p. 319.

MPG. (n.d.). TIMSS III Ergebnisse. Regionale Leistungsvegleiche in Deutschland. Retrieved 2010-04-15, from http://www.timss.mpg.de/

Müller, A. (2003). Dem Wissen auf der Spur. Retrieved 2013-05-23, from http://www.institutbeatenberg.ch/images/publikationen-und-materialien/dossiers/dem_wissen_auf_der_spur.pdf

Müller, P. (2013). Das „Musterländle" auf schulpolitischen Abwegen Retrieved 2013-06-23, from http://bildung-wissen.eu/fachbeitraege/schule-und-unterricht/das-musterlandle-auf-schulp olitischen-abwegen.html

Mwagore, D. (Ed.). (n.d.). *LAND USE IN KENYA The case for a national land use policy* (Vol. 3). Nakuru: Kenya Land Alliance.

National Center for Education Statistics. (2002). *The Nation's Report Card Geography 2001*: US Department of Education.

National Center for Education Statistics. (2011). The Nation's Report Card Science 2009. National Assessment of educational progress at grades 4, 8, and 12. Retrieved 2011-05-04, from http://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf

National Geographic Education Foundation. (2002). National Geographic - Roper 2002 Global Geographic Literacy Survey. Retrieved 2008-09-17, from http://www.nationalgeographic. com/geosurvey2002/download/RoperSurvey.pdf

National Research Council (Ed.). (2006). *Learning to think spatially*. Washington: National Academies Press.

Naumann, S., Volz, D., & Viehrig, K. (2008). 1x1 des GIS. *Praxis Geographie, 38*(7-8), Beilage.

NCGE. (2009). *Geography for life, 2nd edition, national geography standards*: draft version, June 12, 2009.

NCREL. (n.d.). Affective Dimensions of Learning. Retrieved 2014-03-16, from http://www.ncrel .org/sdrs/areas/issues/students/atrisk/at7lk7.htm

Nepal Geographic Information System Society. (2008). Press Release. Retrieved 2008-11-29, from http://negiss.org.np/events.php

Neria, D., Amit, M., Abu-Naja, M., & Abo-Ras, Y. (2008). Ethnic and gender gaps in mathematical self-concept: the case of Bedouin and Jewish students. Retrieved 2010-04-15, from http://74.125.155.132/scholar?q=cache:Q-IJkoK88b0J:scholar.google.com/+%22self-effica cy%22+Arab++high+school+israel&hl=de&as_sdt=2000&as_ylo=2001&as_vis=1

Neumann-Mayer, U. P. (2005). *Der Zugang zu Satellitenbildern in der Orientierungsstufe - Probleme und Möglichkeiten* - Dissertation, Christian-Albrechts-Universität, Kiel.

Newcombe, N. S. (2006). A Plea for Spatial Literacy. *Chronicle of Higher Education, 52*(26).

Niemz, G., & Stoltman, J. P. (1993). InterGeo II: International Geographical Achievement Test. Field Trials Report and Test (Secondary Schools, Grade 8). Retrieved 2009-09-14, from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_ SearchValue_0=ED372985&ERICExtSearch_SearchType_0=no&accno=ED372985

Nobel Foundation. (2004). Wangari Maathai - Biographical. Retrieved 2013-08-12, from http:// www.nobelprize.org/nobel_prizes/peace/laureates/2004/maathai-bio.html

Nölker, D. (1988). Als Tourist in die Dritte Welt. Hilfe oder Hemmnis von Entwicklung? *Praxis Geographie*(3), 10-14.

Nominal measures of correlation: Phi, the Contingency Coefficient, and Cramer's V. (n.d.). Retrieved 2011-05-08, from http://www.harding.edu/sbreezeel/460%20Files/Statbook/ chapter15.pdf

NSSE. (2006). *Technical guide to school and district factors impacting student learning*. Schaumburg: National Study of School Evaluation.

Nusche, D. (2008). What works in migrant education? A review of evidence and policy options. *OECD Education Working Paper*(22).

O'Connor, P. (2005). Geographical Information Systems (GIS) and Issues of Progression in the Learning and Teaching of Geography. Retrieved 2008-09-20, from www.geography.org.uk/ download/GA_PRSSBishopsStortford.doc

Obermaier, G. (1997). *Strukturen und Entwicklung des geographischen Interesses von Gymnasialscülern in der Unterstufe – eine bayernweite Untersuchung*. Lehrstuhl für Didaktik der Geographie der Universität München: Münchner Studien zur Didaktik der Geographie, Band 9.

Obermaier, G. (2002a). *Menschen oder Räume – Was interessiert Mädchen und Jungen im Geographieunterricht?* presented at the 28. Deutscher Schulgeographentag, Wien.

Obermaier, G. (2002b). Umwelt – nein danke? Ein Interessenvergleich zwischen Schülern der deutschen Schulen in Kuala Lumpur und Singapur und Schülern aus Deutschland. *Geographie heute*(202), S. 18-19.

Obermaier, G. (2005). Die Akzeptanz des Internets im Geographieunterricht. Eine empirische Untersuchung. Retrieved 2010-05-10, from http://opus4.kobv.de/opus4-ku-eichstaett/frontdoor/index/index/docId/29

Obermann, H. (Ed.). (2005). *Terra GWG Geographie Wirtschaft 3/4 Gymnasium Baden-Württemberg*. Leipzig: Klett.

Odunga, P., & Folmer, H. (2004). Profiling Tourists for Balanced Utilization of Tourism-Based Resources in Kenya, FEEM Working Paper No. 23.2004 Available from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=504443

OECD. (2000). PISA 2000: student questionnaire. Retrieved 2010-08-18, from http://pisa2000.acer.edu.au/downloads.php

OECD. (2003a). PISA 2003 Student questionnaire. Retrieved 2010-08-18, from http://www.oecd.org/dataoecd/34/7/37617728.pdf

OECD. (2003b). Wo haben Schüler mit Migrationshintergrund die größten Erfolgschancen: Eine vergleichende Analyse von Leistung und Engagement in PISA 2003. Kurzzusammenfassung. Retrieved 2008-08-18, from http://www.pisa.oecd.org/dataoecd/2/57/ 36665235.pdf

OECD. (2005). PISA 2003 Data Analysis Manual: SPSS Users. Retrieved 2011-07-12, from http://www.oecd.org/dataoecd/35/51/35004299.pdf

OECD. (2006a). Assessing Scientific, Reading and Mathematical Literacy. A Framework for PISA 2006. Retrieved 2009-05-10, from http://www.oecd.org/dataoecd/63/35/37464175.pdf

OECD. (2006b). *Where immigrant students succeed - a comparative review of performance and engagement in PISA 2003*. Paris.

OECD. (2007a). *PISA 2006: Naturwissenschaftliche Kompetenzen für die Welt von Morgen. OECD briefing note für Deutschland*. from http://www.oecd.org/germany/39727140.pdf

OECD. (2007b). PISA 2006: Science Competencies for Tomorrow's World, Vol. 1. Retrieved 2010-03-22, from http://www.oecd.org/dataoecd/30/17/39703267.pdf

OECD. (2007c). PISA 2006: Vol. 2 Data/ Données. Retrieved 2010-03-22, from http://www.oecd.org/dataoecd/30/18/39703566.pdf

OECD. (2009). Top of the class - high performers in science in PISA 2006. Retrieved 2010-03-22, from http://www.oecd.org/dataoecd/44/17/42645389.pdf

OECD. (2011). PISA 2009 Results: Students On Line. Digital Technologies and Performance (Volume VI). Retrieved 2014-06-29, from http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2009-results-students-on-line_9789264112995-en

Oldakowski, R. K. (2001). Activities to Develop a Spatial Perspective Among Students in Introductory Geography Courses. *Journal of Geography, 100*(6), 243-250.

Olkun, S. (2003). Making Connections: Improving Spatial Abilities with Engineering Drawing Activities. *International Journal of Mathematics Teaching and Learning*.

Olsen, A. (2001). Using GIS software in school teaching programmes - An intial survey. Retrieved 2007-06-05, from http://egis.eagle.co.nz/schools/Documents/Using%20GIS%20software%20in%20school%20teaching%20programmes.pdf

Ordnance Survey. (2004). The use of GIS in schools – questionnaire results. *Mapping News*(27), 23-24.

Orion, N., & Basis, T. (2008). *Characterization of High School Students' System Thinking Skills in the Context of Earth Systems*. presented at the NARST Annual Conference, Baltimore, USA.

Orion, N., & Cohen, C. (2007). A Design-based Research of an Oceanography Module as a part of the Israeli High School Earth Sciences Program. *Geographie und Ihre Didaktik*(4), 246-259.

Orr, D. (1994). *Earth in Mind: On Education, Environment and the Human Prospect*. Washington DC: Island Press.

Ossimitz, G. (1996). *Das Projekt "Entwicklung vernetzten Denkens". Erweiterter Endbericht*: Universität Klagenfurt.

Ossimitz, G. (2000). *Entwicklung systemischen Denkens*. Wien: Profil.

Ossimitz, G. (2001). *Stock-Flow-Thinking and Reading Stock-Flow-Related-Graphs: An Empirical Investigation in Dynamic Thinking Abilities*. presented at the 19th International Conference of the System Dynamics Society, July 23-27, 2001.

Ossimitz, G., Kotzent-Pietschnig, U., Kreisler, B., Waiguny, M., & Zoltan, M. (2001). *Endbericht zum Projekt. Unterscheidung von Bestands- und Flussgrößen*: Universität Klagenfurt.

Otto, K.-H., Mönter, L., & Hof, S. (2011). (Keine) Experimente wagen? In C. Meyer, R. Henry & G. Stöber (Eds.), *Geographische Bildung. Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 98-113). Braunschweig: Westermann.

Palmer, A. M., Palmer, R., Malone, L., & Voigt, C. L. (2008a). *Mapping Our World Using GIS. Level 2*. Redlands: ESRI Press.

Palmer, A. M., Palmer, R., Malone, L., & Voigt, C. L. (2008b). *Mapping Our World Using GIS. Student Workbook. Level 2*. Redlands: ESRI Press.

Palmer, R., Palmer, A. M., Malone, L., & Voigt, C. L. (2008c). *Analyzing Our World Using GIS. Level 3.* Redlands: ESRI Press.

Palmer, R., Palmer, A. M., Malone, L., & Voigt, C. L. (2008d). *Analyzing Our World Using GIS. Level 3. Student Workbook*. Redlands: ESRI Press.

Pastorelli, C., Caprara, G. V., Barbaranelli, C., Rola, J., Rozsa, S., & Bandura, A. (2001). The structure of children's perceived self-efficacy: a cross-national study. *European Journal of Psychological Assessment, 17*(2), 87-97.

Patterson, M. W., Reeve, K., & Page, D. (2003). Integrating Geographic Information Systems into the Secondary Curricula. *Journal of Geography, 102*, 275-281.

Peer, A., Gortan, E., Haider, R., Painer, C., Bauer, E., Rosmarin, W., et al. (2008). Erproben einer neuen Didaktik für die Einführung der Proportionen. Retrieved 2012-11-04, from https://www.imst.ac.at/imst-wiki/images/1/1c/997_Langfassung_Peer.pdf

Penner, D. E. (2000). Explaining Systems: Investigating Middle School Students' Understanding of Emergent Phenomena. *Journal of Research in Science Teaching, 37*(8), 784-806.

Petersen, F. C. (2007). *Neurodidaktik- eine neue Didaktik? Eine Untersuchung zu einer neurowissenschaftlich geleiteten Didaktik sowie ein vergleichender Exkurs zum Behaviorismus*. Diplom Thesis, Universität Flensburg, Flensburg.

Peterßen, W. H. (2000). *Handbuch Unterrichtsplanung: Grundfragen, Modelle, Stufen, Dimensionen*. München.

Pfeiffer, C., & Pfeiffer, R. (2011). Computerspielsucht. Was haben Online-Spiele mit Glücksspielautomaten gemeinsam? Wie tragen sie dazu bei, dass vor allem mannliche [sic!] Jugendliche in exzessives oder gar suchtartiges Spielen geraten? Retrieved 2013-11-12, from http://medienverleihstelle.rpz-basel.ch/fileadmin/daten/medien/pdf/Computerspielsucht_-_der_grosse_Leistungskiller.6-10-11doc.pdf

Phan, T. T. H. (2007). *Testing levels of competencies in biological experimentation.* Dissertation, Christian-Albrechts-University, Kiel.

Philologenverband Baden-Württemberg. (2008). Kritik an G8-Kürzungsvorschlägen von Ministerpräsident Oettinger: Landesregierung muss deutlich mehr in Bildung investieren. Retrieved 2010-03-22, from http://www.phv-bw.de/Veroeffentlichung/Publikationen/GBW_2008_03/02-G-8-Schnellschuesse.html

Phoenix, M. (2004). GIS education: A global summary. Retrieved 2008-11-29, from http://www.eugises.eu/proceedings2004/files/01%20-%20EUGISES04_paper_phoenix.pdf

Pieter, A. (2003). *Selbstbestimmtes Lernen in der Schule, Erfassung der subjektiven Kompetenz zum selbstbestimmten Lernen.* Dissertation, Universität des Saarlandes.

PISA-Konsortium Deutschland (Ed.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.

Pönitz, S. (2008). *Das Unterrichten von GIS an Baden-Württembergischen Gymnasien am Beispiel des Projektes "Umweltatlas".* Hochschule für Wirtschaft und Umwelt Nürtingen-Geislingen.

Popp, M. (2007, 08.05.2007). US-Schulen schwören Computern ab. *Spiegel Online*.

Porsch, O. (2005). Thematische Karten mit GIS gestalten, Eine problemorientierte Einführung in SchulGIS. *Geographie heute, 26*(233), 33-35.

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students. Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted child quarterly, 52*(2), 146-159.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., et al. (2008). PISA 2006 in Deutschland - Die Kompetenzen der Jugendlichen im dritten Ländervergleich. Retrieved 2010-01-05, from http://www.wir-wollen-lernen.de/resources/PISA-E_2006_vollst_Bericht_HH.pdf

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (2004). (Eds.) PISA Konsortium Deutschland: PISA 2003. Ergebnisse des zweiten internationalen Vergleichs. Zusammenfassung. Retrieved 2007-07-13, from http://pisa.ipn.uni-kiel.de/ Ergebnisse_PISA_2003.pdf

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (2005). PISA 2003: Ergebnisse des zweiten Ländervergleichs. Zusammenfassung. Retrieved 2008-10-07, from http://pisa.ipn.uni-kiel.de/PISA2003_E_Zusammenfassung.pdf

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (Eds.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres. Zusammenfassung*. Kiel: IPN.

Prenzel, M., & Deutsches Pisakonsortium. (2005). *PISA 2003: Der zweite Vergleich der Länder in Deutschland - was wissen und können Jugendliche?*: Waxmann.

Püschel, L. (2004). DierckeGIS am Beispiel Landwirtschaft – Eine methodische Einstiegshilfe für Anfänger. *Praxis Geographie, 34*(2), 16-20.

Püschel, L. (2005). Internetbasierte GIS-Anwendungen für den Einstieg in die GIS-Arbeit. *Geographie heute, 26*(233), 18-21.

Püschel, L. (2007). Vom Web-GIS zum Desktop-GIS - ein didaktisch-methodisches Konzept zur Einführung von Geographischen Informationssystemen an Schulen. In T. Jekel, A. Koller & J. Strobl (Eds.), *Lernen mit Geoinformation II* (pp. 138-148). Heidelberg: Wichmann.

Püschel, L. (2011). *Diercke. Lernen mit GIS. Vom Web-GIS zum Desktop-GIS*. Braunschweig: Westermann.

Püschel, L., Hofmann, K., & Hermann, N. (2007). *Blickpunkt WebGIS. Der Einstieg für Schulen in Geographische Informationssysteme (GIS)*. Koblenz: Landesmedienzentrum Rheinland-Pfalz.

Püschel, L., & Schäfer, D. (2004). GIS im Internet, Das Unterrichtsbeispiel „Das Klima von Deutschland". *Praxis Geographie, 34*(2), 39-43.

Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences*(40), 609-619.

Quaiser-Pohl, C., & Lehmann, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology, 72*, 245-260.

Radisch, F. (2010). Einführung in die Mehrebenenanalys mit HLM 6. AEPF/ Universität Jena. Retrieved 2013-05-11, from http://www.aepf.uni-jena.de/aepfmedia/pdfs/Foliensatz+Jena.pdf

Raia, F. (2005). Students' Understanding of Complex Dynamic Systems. *Journal of Geoscience Education, 53*(3), 297-308.

Rauh, J. (2006). United or Divded? Die Welt im Informationszeitalter. *Geographische Rundschau, 58*(7/8), 4-11.

Rebich, S., & Gautier, C. (2005). Concept Mapping to Reveal Prior Knowledge and Conceptual Change in a Mock Summit Course on Global Climate Change. *Journal of Geoscience Education, 53*(4), 355-365.

Reinfried, S. (2006). Interessen, Vorwissen, Fähigkeiten und Einstellungen von Schülerinnen und Schülern berücksichtigen. In H. Haubrich (Ed.), *Geographie unterrichten lernen - Die neue Didaktik der Geographie konkret* (pp. 49-78). München: Oldenbourg.

Reitz, V. (2005). Nutzungskonflikte im Nationalpark Schleswig-Holsteinisches Wattenmeer, Inhaltliches Lernen mit SchulGIS. *Geographie heute, 26*(233), 41-43.

Rempfler, A. (1998). *Das Geoökosystem und seine schuldidaktische Aufarbeitung*. Basel: Wepf & Co.

Rempfler, A. (1999). Geoökologie und Systemdenken. *Geographie und ihre Didaktik, 27*(4), 173-191.

Rempfler, A., & Uphues, R. (2010a). Sozialökologisches Systemverständnis: Grundlage für die Modellierung von geographischer Systemkompetenz. *Geographie und ihre Didaktik*(4), 205-217.

Rempfler, A., & Uphues, R. (2010b). *Systemkompetenz im Geographieunterricht – Konstrukt, Diagnostik, Förderung*. presented at the Joint Symposium of the GEI and the HGD, March 18-19, 2010, Braunschweig, Germany.

Rempfler, A., & Uphues, R. (2011a). Für ein adäquates Verständnis von Geosystemen. *Geographie und Schule, 33*(189), 4-10.

Rempfler, A., & Uphues, R. (2011b). Systemkompetenz im Geographieunterricht – Die Entwicklung eines Kompetenzmodells. In C. Meyer, R. Henry & G. Stöber (Eds.), *Geographische Bildung. Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 36-48). Braunschweig: Westermann.

Rempfler, A., & Uphues, R. (2011c). Systemkompetenz und ihre Förderung im Geographieunterricht. *Geographie und Schule, 33*(189), 22-26,31-33.

Rempfler, A., & Uphues, R. (2012). System competence in geography education. Development of competence models, diagnosing pupils' achievement. *European Journal of Geography, 3*(1), 6-22.

Rheinberg, F., Vollmeyer, R., & Burns, B., D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen (Langversion). Retrieved 2013-07-07, from http://www.psych.uni-potsdam.de/people/rheinberg/messverfahren/FAMLangfassung.pdf

Rheinberg, F., & Wendland, M. (2003). Veränderung der Lernmotivation in Mathematik und Physik: eine Komponentenanalyse und der Einfluss elterlicher sowie schulischer Kontextfaktoren. Abschlussbericht (2003) zum DFG-Projekt.

Rhode-Jüchtern, T. (2009). *Eckpunkte einer modernen Geographiedidaktik Hintergrundbegriffe und Denkfiguren (mit je einem Glossar von Volker Schmidtke und Karen Krösch)* (1. Aufl. ed.). Seelze: Kallmeyer.

Rider, N., & Colmar, S. (2005). *Reading achievement and reading self-concept in year 3 children*. presented at the Australian Association for Research in Education Conferencem, Parramatta, 27 Nov - 1 Dec 2005. Retrieved from www.aare.edu.au/05pap/col05347.pd

Rieß, W., & Mischo, C. (2008a). Entwicklung und erste Validierung eines Fragebogens zur Erfassung des systemischen Denkens in nachhaltigkeitsrelevanten Kontexten. In I. Bormann & G. De Haan (Eds.), *Kompetenzen der Bildung für nachhaltige Entwicklung. Operationalisierung, Messung, Rahmenbedingungen, Befunde.* (pp. 215-232). Wiesbaden: Verlag für Sozialwissenschaften.

Rieß, W., & Mischo, C. (2008b). Wirkungen variierten Unterrichts auf systemisches Denken. In U. Frischknecht-Tobler, U. Nagel & H. Seybold (Eds.), *Systemdenken. Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (pp. 135-147). Zürich: pestalozzianum.

Rinschede, G. (2003). *Geographiedidaktik*. Paderborn.

Rinschede, G. (2005). *Geographiedidaktik*. Paderborn.

Roberts II, T. G. (2003). *The influence of student characteristics on achievement and attitudes when an illustrated web lecture is used in an online learning environment.* University of Florida.

Robertson, M. M., Althoff, R. R., Hafez, A., & Pauls, D. (2008). Principal components analysis of a large cohort with Tourette syndrome. *The British Journal of Psychiatry, 193*(1), 31-36.

Robinsohn, S. B. (1967). *Bildungsreform als Revision des Curriculum*. Berlin: Luchterhand.

Rollett, B. (2011). Null Lust auf gar nichts! Geschlechtsspezifische Unterschiede in der Leistungsverweigerung, ihre Ursachen und Interventionsmöglichkeiten. Retrieved 2013-05-23, from http://www.fh-ooe.at/fileadmin/fileSystem/Konferenzen/Gendertagung/Gendertagung_Linz_2011.pdf

Roper Public Affairs, & National Geographic Education Foundation. (2006). *Final Report National Geographic-Roper Public Affairs 2006 Geographic Literacy Study*. Washington D.C.

Rost, D. H. (2005a). *Interpreation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung*. Weinheim: Beltz.

Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Hans Huber.

Rost, J. (2005b). Messung von Kompetenzen Globalen Lernens. *Zeitschrift für ZEP, 2005*(2), 14-18.

Rost, J., Lauströer, A., & Raack, N. (2003). Kompetenzmodelle einer Bildung für Nachhaltigkeit. *Praxis der Naturwissenschaften - Chemie in der Schule, 52*(8), 10-15.

RTVE.ES. (2014). Gallardón anuncia que la nacionalidad requerirá un examen de español en el Instituto Cervantes. Retrieved 2014-06-23, from http://www.rtve.es/noticias/20140422/gallardon-anuncia-nacionalidad-requerira-examen-espanol-instituto-cervantes/925002.shtml

Ruiz-Primo, M. A. (2004). Examining concept maps as an assessment tool. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology Proc. of the First Int. Conference on Concept Mapping*. Pamplona.

Rumpf, H.-J., Meyer, C., Kreuzer, A., & John, U. (2011). Prävalenz der Internetabhängigkeit (PINTA). Bericht an das Bundesministerium für Gesundheit. Retrieved 2013-11-12, from http://drogenbeauftragte.de/fileadmin/dateien-dba/DrogenundSucht/Computerspiele_Internetsucht/Downloads/PINTA-Bericht-Endfassung_280611.pdf

Ruslanguage School Moscow. (2014). Certificates and TORFL. Retrieved 2014-06-23, from http://www.russian-moscow.com/study/certificates/

Russell, T. L. (2010). no significant difference. Retrieved 2013-06-01, from http://www.nosignificantdifference.org

Rychen, D. S., & Salganik, L. H. (2003). Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo) Summary of the final report "Key Competencies for a Successful Life and a Well-Functioning Society". Retrieved 2007-11-27, from www.portal-stat.admin.ch/deseco/deseco_finalreport_summary.pdf

Sacerdote, B. (2001). Peer effects with random assigment: results for dartmouth roommates. *The quarterly journal of economics*(May), 681-704.

Sander, M. (2002). Einsatzmöglichkeiten für ein Geographisches Informationssystem im Geographieunterricht des Gymnasiums - Beispiele und praktische Hinweise. *Geographie und Schule*(139), 24-25.

Sapp, M. (2006). *Basic psychological measurement, research designs, and statstics without math*. Springfield: Charles C Thomas.

Scanlan, C. L. (n.d.). Introduction to Nonparametric Statistics. Retrieved 2011-05-08, from http://www.umdnj.edu/idsweb/idst6000/nonparametric_analysis.pdf

Schäfer, D. (2002). *Geographische Informationssysteme (GIS) im Erdkundeunterricht*. presented at the 9. Deutschsprachige ESRI Anwenderkonferenz, Konferenz-CD.

Schäfer, D. (2007). *Skalierbarer Einsatz von Geographischen Informationssystemen (GIS) in Schulen.* Paper presented at the 3. GIS Ausbildungstagung, 07./08. Juni 2007, Konferenz-CD, Potsdam.

Schäfer, F. (2005). Arbeiten mit dem GIS-Nationalatlas USA, Eine GIS-gestützte Einführung in die Landwirtschaft der USA. *Geographie heute, 26*(233), 22-25.

Schallhorn, E. (2004a). Ziele des Geographieunterrichts. In E. Schallhorn (Ed.), *Erdkundedidaktik - Praxishandbuch für die Sekundarstufe I und II* (pp. 24-33). Berlin: Cornelsen Scriptor.

Schallhorn, E. (2007). *Erdkunde Methodik. Handbuch für die Sekundarstufe I und II*. Berlin: Cornelsen.

Schallhorn, E. (Ed.). (2004b). *Erdkunde Didaktik. Praxishandbuch für die Sekundarstufe I und II*. Berlin.

Schecker, H., Klieme, E., Niedderer, H., Ebach, J., & Gerdes, J. (1999). *Abschlussbericht zum DFG Projekt Physiklernen mit Modellbildungssystemen. Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildungssysteme*. Universität Bremen: Institut für Bildungsforschung (Bonn).

Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften, 12*, 45-66.

Schermelleh-Engel, K., & Werner, C. (2007). Uni- und multivariate Varianzanalyse. Retrieved 2011-05-10, from http://user.uni-frankfurt.de/~cswerner/multivariate/anova_manova.pdf

Schleicher, Y. (2002a). Mit Interesse im Internet surfen. Welche geographen Websites interessieren Schüler? – Ergebnisse einer empirischen Studie. *Geographie heute*(202), S. 16-17.

Schleicher, Y. (2002b). *Nutzen Schüler geographische Websites? Eine empirische Studie*. Nürnberg: HGD.

Schleicher, Y. (2004a). GIS - Geographische Informationssysteme. In Y. Schleicher (Ed.), *Computer, Internet & Co. im Erdkundeunterricht* (pp. 188-190). Berlin.

Schleicher, Y. (2004b). GIS im Geographieunterricht - Machen wir uns den Einstieg zu schwer? *Der Bayrische Schulgeograph, 25*(55), 12-13.

Schleicher, Y. (2006). Arbeiten und Präsentieren mit Geographischen Informationssystemen (GIS). In H. Haubrich (Ed.), *Geographie unterrichten lernen – die neue Didaktik der Geographie konkret* (pp. 218-219). München.

Schmenk. B. (2004). Interkulturelles Lernen versus Autonomie? In W. Börner & K. Vogel (Eds.), *Emotion und Kognition im Fremdsprachenunterricht* (pp. 66-86). Tübingen.

Schmitz, A. (2006). *Interessen- und Wissensentwicklung bei Schülerinnen und Schülern der Sek II in außerschulischer Lernumgebung am Beispiel von NaT-Working „Meeresforschung".* Dissertation, Christian-Albrechts-Universität, Kiel.

Schostak, J. F. (2002). *Understanding, designing and conducting qualitative research in education framing the project*. Buckingham: Open University Press.

Schreiber, J. R. (2005). Kompetenzen und Konvergenzen. Globales Lernen im Rahmen der UN-Dekade „Bildung für nachhaltige Entwicklung". *ZEP, 28*(2), 19-25.

Schubert, J. C., & Uphues, R. (2008). Kumulatives Lernen mit Geoinformation. Überlegungen zu einem GI(S)-Kompetenzentwicklungsmodell. In T. Jekel, A. Koller & K. Donert (Eds.), *Lernen mit Geoinformation III - Learning with Geoinformation III* (pp. 49-59). Heidelberg: Wichmann.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah: Lawrence Erlbaum Associates.

Schwab, A., & Kussmaul, G. (2005). Thematische Karten mit GIS erstellen und auswerten, Anregungen für eine Untersuchung der Bevölkerungsentwicklung im Nahraum (Beispiel Bodensee-Oberschwaben). *Geographie heute, 26*(229), 38-45.

Schwab, J. (2007). The analysis of covariance. Retrieved 2011-05-10, from http://www.utexas.edu/courses/schwab/sw388r7_spring_2007/SolvingProblemsInSPSS/Solving%20Analysis%20of%20Covariance%20Problems.pdf

Schwarz, J.-A. (2005). *Kommunikation von Geoinformation: Mehrwert durch den Einsatz Neuer Medien? Eine Untersuchung am Beispiel internetbasierter Lernangebote.* Dissertation, Universität Potsdam.

Self, C. M., & Golledge, R. G. (1994). Sex-related Differences in Spatial Ability: What Every Geography Educator Should Know. *Journal of Geography, 93*(5), 234-243.

Sell, K. S., Herbert, B. E., Stuessy, C. L., & Schielack, J. (2006). Supporting Student Conceptual Model Development of Complex Earth Systems Through the Use of Multiple Representations and Inquiry. *Journal of Geoscience Education, 54*(3), 396-407.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*(1), 159-189.

Sharpe, B., & Huynh, N. T. (2008). Spatial reasoning abilities and effective GIS problem-solving: an empirical study. In A. Demirci, M. Karakuyu, M. A. McAdams, S. Incekara & A. Karaburun (Eds.), *5th international conference on geographic information systems. Proceedings* (Vol. 1, pp. 131-138). Istanbul: Fatih University.

Shin, E.-k. (2006). Using Geographic Information System (GIS) to Improve Fourth Graders' Geographic Content Knowledge and Map Skills. *Journal of Geography, 105*(3), 109-120.

Sibley, D. F., Anderson, C. W., Heidemann, M., Merill, J. E., Parker, J. M., & Szymanski, D. W. (2007). Box Diagrams to Assess Students' Systems Thinking about the Rock, Water and Carbon Cycles. *Journal of geoscience education, 55*(2), 138-146.

Siegmund, A. (2001). Die GIS-Layertechnik. Von der Karte zum selbst gefertigten "Hardware-GIS". *Geographie heute, 22*(195), 22-26.

Siegmund, A. (2002). Neue und traditionelle Medien im Geographieunterricht, Medienverbund als Chance für handlungsorientiertes Lernen. *Praxis Geographie, 32*(6), 4-8.

Siegmund, A. (2008). Die Klimakarte nach Siegmund/ Frankenberg im Baukastensystem. Retrieved 2012-06-01, from http://www.diercke.de/bilder/omeda/siegmund_sek1_02_2008.pdf

Siegmund, A. (2010). *Satellitenbilder im Unterricht – eine Ländervergleichsstudie zur Ableitung fernerkundungsdidaktischer Grundsätze.* Dissertation, Pädagogische Hochschule Heidelberg, Heidelberg.

Siegmund, A., Funke, J., Viehrig, K., Wüstenberg, S., & Greiff, S. (2012). Theoriegeleitete Erhebung von Kompetenzstufen im Rahmen probabilistischer Messmodelle – ein Beitrag zum Aufbau eines Heidelberger Inventars Geographischer Systemkompetenz (HEIGIS) – Abschlussbericht. Unpublished Final Project Report Submitted to the DFG.

Siegmund, A., & Naumann, S. (2009). GIS in der Schule. Potenziale für den Geographieunterricht von heute. *Praxis Geographie*(2), 4-8.

Siegmund, A., & Viehrig, K. (2012). Internationale Vernetzung. In J. B. Haversath (Ed.), *Geographiedidaktik. Theorie – Themen – Forschung* (pp. 56-63): Westermann.

Siegmund, A., Viehrig, K., & Volz, D. (2008). GIS@school - new didactical aspects of using GIS in geography education. In K. Donert (Ed.), *EUC'07 HERODOT Proceedings, ESRI European User Conference 2007: Stockholm, Sweden, 25-27 September 2007*. ESRI-HERODOT Publications.

Siegmund, A., Viehrig, K., & Volz, D. (2009). Mit GIS geographische Erkenntnisse gewinnen. Konzept eines Kompetenzmodells. *Praxis Geographie, 39*(2), 10-11.

Siegmund, A., Volz, D., & Viehrig, K. (2007). GIS in the classroom - challenges and chances for geography teachers in Germany. *Papers of the HERODOT working conference, Stockholm, Thematic Pillar 4: 'Employability'*. Retrieved from http://www.herodot.net/conferences/stockholm/HERODOT-Stockholm2.html#pres

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha *Psychometrika, 74*(1), 107-120.

SIL International. Retrieved 2012-01-11, from www.ethnologue.com

Simmons, M. E., Wu, X. B., Knight, S. L., & Lopez, R. R. (2008). Assessing the Influence of Field- and GIS-based Inquiry on Student Attitude and Conceptual Knowledge in an Undergraduate Ecology Lab. *CBE Life Sciences Education, 7*(3), 338-345.

Simon, S. (2010). What's the difference between regression and ANOVA? Retrieved 2011-05-10, from http://www.pmean.com/08/RegressionAndAnova.html

Sinton, D. S., & Bednarz, S. W. (2007). About that G in GIS. In D. S. Sinton & J. J. Lund (Eds.), *Understanding Place. GIS and Mapping Across the Curriculum* (pp. 19-33). Redlands: ESRI Press.

Sinton, D. S., & Lund, J. J. (Eds.). (2007). *Understanding Place. GIS and Mapping Across the Curriculum*. Redlands: ESRI Press.

Slavin, R. E. (2008). What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher, 37*(1), 5-14.

Smith, T. W., Gordon, B., Colby, S. A., & Wang, j. (2005). *An Examination of the Relationship Between Depth of Student Learning and National Board Certification Status*: Office for Research and Teaching Appalachian State University.

Smithson, P., Addison, K., & Atkinson, K. (2002). *Fundamentals of the physical environment*. London: Routledge.

Sommer, C. (2005). *Untersuchung der Systemkompetenz von Grundschülern im Bereich Biologie.* Dissertation, Christian-Albrechts-Universität, Kiel.

Songer, L. C. (2010). Using Web-Based GIS in Introductory Human Geography. *Journal of Geography in Higher Education, 34*(3), 401-417.

SPSSX Discussion. (2009). cronbach alpha for binary responses. Retrieved 2013-08-22, from http://spssx-discussion.1045642.n5.nabble.com/cronbach-alpha-for-binary-responses-td1086058.html

SSI Central. (n.d.-a). Full ML vs. Restricted ML. Retrieved 2013-05-11, from http://www.ssicentral.com/hlm/help6/faq/Full_ML_vs_Restricted_ML.pdf

SSI Central. (n.d.-b). Proportion Variance Explained: Two-level models. Retrieved 2013-05-11, from http://www.ssicentral.com/hlm/help6/faq/Proportion_Variance_Explained__Two-level_models.pdf

SSI Central. (n.d.-c). Use of proportion variance explained statistics. Retrieved 2014-06-17, from http://www.ssicentral.com/hlm/help6/faq/Use_of_proportion_variance_explained_statistics.pdf

Statistical Innovations. (2005). Tutorial #7A: Latent Class Growth Model (#seizures). Retrieved 2011-10-14, from http://statisticalinnovations.com/products/LGtutorial7A.pdf

Statistisches Bundesamt. (2009). *BILDUNGS-FINANZBERICHT 2009. Im Auftrag des Bundesministeriums für Bildung und Forschung und der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland*. Wiesbaden.

Stave, K., & Hopper, M. (2007). What constitutes systems thinking? A proposed taxonomy. Retrieved 2010-03-29, from http://www.systemdynamics.org/conferences/2007/proceed/papers/ STAVE210.pdf

Steinhardt, U., Blumenstein, O., & Barsch, H. (2005). *Lehrbuch der Landschaftsökologie*. Heidelberg: Elsevier Spektrum Akademischer Verlag.

StN. (2010). Landeselternbeiratsvorsitzende: Staab und Wiegert treten zurück. Retrieved 2010-03-22, from http://content.stuttgarter-nachrichten.de/stn/page/2361820_0_9223_-persoenliche-erklaerung-landeselternbeiratsvorsitzende-staab-und-wiegert-treten-zurueck.html

Storie, C. D. (2000). *Assessing the role of geographical information systems (GIS) in the classroom.* MA thesis, Wilfrid Laurier University.

Stracke, I. (2004). *Einsatz computerbasierter Concept Maps zur Wissensdiagnose in der Chemie. Empirische Untersuchungen am Beispiel des Chemischen Gleichgewichts*. Münster.

Strahler, A. H., & Strahler, A. (2002). *Physical geography science and systems of the human environment* (2. ed. ed.). New York: Wiley.

Strong, S., & Smith, R. (2001). Spatial Visualization: Fundamentals and Trends in Engineering Graphics. *Journal of Industrial Technology, 18*(1), 2-6.

Sturm, B. (1981). Ferntourismus in Kenia. *Praxis Geographie, 11*(10), 396-405.

Stutzer, E. (2005). Bildungsintegration von Migranten. *Statistisches Monatsheft Baden-Württemberg*(9), 3-8.

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub Dynamics: Initial Results of a Systems Thinking Inventory. *System Dynamics Review, 16*(4), 249-286.

Sword, L. (n.d.). I Think in Pictures, You Teach in Words: The Gifted Visual Spatial Learner. Retrieved 2010-08-24, from http://www.nswagtc.org.au/ozgifted/conferences/Sword VisualSpatialChildren.html

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson.

Taber, K. S. (2001). Building the structural concepts of chemistry: some considerations from educational research. *Chemistry Education: Research and Practice in Europe, 2*(2), 123-158.

Taylor-Ward, C. (2003). *The Relationship between Elementary School Foreign Language Study in Grades Three through Five and Academic Achievement on the Iowa Tests of Basic Skills (ITBS) and the Fourth-grade Louisiana Educational Assessment Program for the 21st Century (LEAP 21) Test.* Dissertation, Louisiana State University and Agricultural and Mechanical College.

Thompson, K., & Reimann, P. (2007). Do school students learn more about the environment from a system dynamics model by themselves or with a partner? *Conference Proceedings. The 2007 International Conference of the System Dynamics Society and 50th Anniversary Celebration.July 29 – August 2, 2007, Boston, Massachusetts, USA*.

Thyne, J. (2005). Geographic information systems in schools. *New Zealand Geographer, 61*(1), 51-52.

Tillmann, A. (2006). *Konzeption und Entwicklung webbasierter multimedialer Lehr-/Lernangebote in der Geographie. Exemplarische Umsetzung anhand geomorphologischer Prozesse und ihrer Modellierung.* Dissertation, Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Titze, C., Heil, M., & Jansen, P. (2008). Gender Differences in the Mental Rotations Test (MRT) Are Not Due to Task Complexity. *Journal of Individual Differences, 29*(3), 130-133.

Tolvanen, S. (2008). GIS education in Finnish Upper Secondary Schools. Retrieved 2010-09-14, from http://gis.esri.com/library/userconf/educ08/educ/papers/pap_1902.pdf

Treier, R. (2006). *Geografische Informationssysteme (GIS); Grundlagen und Übungsaufgaben für die Sekundarstufe II* (1. Aufl. ed.). Bern: hep.

Trichy, H. (1978). *Traumland Kenia*. Innsbruck: Pinguin-Verlag.

Tschirner, S. (2009). GIS in deutschen Klassenzimmern - von der Wandkarte zu openstreetmap. In T. Jekel, A. Koller & K. Donert (Eds.), *Learning with Geoinformation IV - Lernen mit Geoinformation IV* (pp. 32-38). Heidelberg: Wichmann.

Turconi, G., Celsa, M., Rezzani, C., Biino, G., Sartirana, M. A., & Roggi, C. (2003). Reliability of a dietary questionnaire on food habits, eating behaviour and nutritional knowledge of adolescents. *European Journal of Clinical Nutrition, 57*(753-763).

UCLA Academic Technology Services Statistical Consulting Group. (n.d.). What does Conbach's alpha mean? Retrieved 2011-05-05, from http://www.ats.ucla.edu/stat/spss/faq/ alpha.html

UCLA Statistical Consulting Group. (n.d.). Statistical Computing Seminar: Introduction to Multi-level Modeling Using HLM. Retrieved 2013-05-11, from http://www.ats.ucla.edu/stat/hlm/ seminars/hlm_mlm/608/mlm_hlm_seminar_v608.htm

Uljens, M. (1997). *School Didactics and Learning: A School Didactic Model Framing an Analysis of Pedagogical Implications of Learning Theory*. Hove: Psychology Press.

Universität Erfurt. (2014). Studium im Ausland. Bewerbung, Bewerbungsunterlagen, Sprachkenntnisse und Termine. Retrieved 2014-06-23, from http://www.uni-erfurt.de/international/ outgoing/studium/erasmus/termine/

Universität Heidelberg. (2014). M.A. Konferenzdolmetschen. Retrieved 2014-06-23, from http:// www.uni-heidelberg.de/fakultaeten/neuphil/iask/sued/studium/studiengaenge/ma_konferenz.html

Universität Heidelberg. (n.d.). Süddeutsche Fillia des Halleschen Zertifizierungszentrums für Russisch. Tests online. Elementarstufe. Retrieved 2014-06-23, from http://www.slav.uni-heidelberg.de/trki/was_muss_man_koennen/tests/tests_online/elementarstufe.html

University of Wisconsin Oshkosh. (n.d.). Item Discrimination I. Retrieved 2011-07-03, from http://www.uwosh.edu/testing/facultyinfo/itemdiscrimone.php

Unterthurner, H. (2004). GIS-Projekt „Hotelführer Meran" – Eine Kooperation mit Partnern außerhalb der Schule. *Praxis Geographie, 34*(2), 28-31.

UNWTO. (n.d.). *Tourism Highlights. 2008 Edition.*: UNWTO Publications Department.

Uphues, R. (2007). *Die Globalisierung aus der Perspektive Jugendlicher theoretische Grundlagen und empirische Untersuchungen*. Nürnberg: Selbstverl. d. Hochschulverbandes für Geographie und ihre Didaktik.

Uwah, C. J., McMahon, G. H., & Furlow, C. F. (2008). School Belonging, Educational Aspirations, and Academic Self-Efficacy Among African American Male High School Students: Implications for School Counselors. *Professional School Counseling, 11*(5), 296-305.

Valentine, J., & Cooper, H. (2003). Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes. Retrieved 2011-05-05, from http://ies.ed.gov/ncee/ wwc/pdf/essig.pdf

Vanasupaa, L., Rogers, E., & Chen, K. (2008). *Work in Progress: How Do We Teach and Measure Systems Thinking?* presented at the 38th ASEE/IEEE Frontiers in Education Conference, October 22 – 25, 2008.

VDI. (2007). *Bildungsstandards Technik für den Mittleren Bildungsabschluss*. Düsseldorf.

Verma, G. K., & Mallick, K. (1999). *Researching Education - Perspectives and Techniques*. London: Falmer Press.

Vermunt, J. K., & Magidson, J. (2005). *Technical Guide for Latent Gold 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Vettiger, B., Boller, F., Golay, D., Jetzer, A., Klopfstein, U., Rempfler, A., et al. (2006). *Basismodule Geografie S1*: Lehrmittelverlag des Kantons Zürich.

VGD. (2007). *Bildungsstandards Geschichte*. Schwalbach: Wochenschau.

Viehrig, K. (2009). "Wie ein Elefant im …". *Praxis Geographie Extra. Praxis Blätter 7. - 8. Klasse*, 34.

Viehrig, K. (2013). Kompetenzmodellierung. In D. Böhn & G. Obermaier (Eds.), *Wörterbuch der Geographiedidaktik. Begriffe von A-Z* (pp. 157-158). Braunschweig: Westermann.

Viehrig, K., Greiff, S., Siegmund, A., & Funke, J. (2011). Geographische Kompetenzen fördern – Erfassung der Geographischen Systemkompetenz als Grundlage zur Bewertung der Kompe-

tenzentwicklung. In C. Meyer, R. Henrÿ & G. Stöber (Eds.), *Geographische Bildung: Kompetenzen in didaktischer Forschung und Schulpraxis* (pp. 49-57). Braunschweig: Westermann.

Viehrig, K., & Siegmund, A. (2012). Germany: Diverse GIS Implementations within a Diverse Educational Landscape. In A. Milson, A. Demirci & J. J. Kerski (Eds.), *International Perspectives on Teaching and Learning with GIS in Secondary Schools* (pp. 107-113). Dordrecht: Springer.

Viehrig, K., Siegmund, A., Wüstenberg, S., Greiff, S., & Funke, J. (2012). Systemisches und räumliches Denken in der geographischen Bildung – Erste Ergebnisse zur Überprüfung eines Modells der Geographischen Systemkompetenz. In A. Hüttermann, P. Kirchner, S. Schuler & K. Drieling (Eds.), *Räumliche Orientierung: Räumliche Orientierung, Karten und Geoinformation im Unterricht* (pp. 95-102). Braunschweig: Westermann.

Viehrig, K., & Volz, D. (2009). Tourismus in Kenia. Chancen und Probleme im Spannungsfeld komplexer Mensch-Umwelt-Beziehungen. *Praxis Geographie, 39*(2), 18-21.

Viehrig, K., & Volz, D. (2013). Raumverhaltenskompetenz. In D. Böhn & G. Obermaier (Eds.), *Wörterbuch der Geographiedidaktik. Begriffe von A-Z* (pp. 230-231). Braunschweig: Westermann.

Viehrig, K., Volz, D., & Siegmund, A. (2008). A question of objective: implementing GIS-use in secondary schools. In A. Demirci, M. Karakuyu, M. A. McAdams, S. Incekara & A. Karaburun (Eds.), *5th International Conference on Geographic Information Systems - Proceedings*. Istanbul: Fatih University Department of Geography.

Vogt, H. (2007). Theorie des Interesses und des Nicht-Interesses. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (pp. 9-20). Berlin: Springer.

Volz, D., & Siegmund, A. (2013). GeoÖko-Labor/ ReKli:B. Umweltbildung der Abteilung Geographie an der Pädagogischen Hochschule Heidelberg erfolgreich gestartet. *HGG Journal*(27), 120-122.

Volz, D., Viehrig, K., & Siegmund, A. (2008). GIS as a means for competence development – Questions for an integrated GIS Didactics. In T. Jekel, A. Koller & K. Donert (Eds.), *Lernen mit Geoinformation III* (pp. 42-48). Heidelberg: Wichmann.

Volz, D., Viehrig, K., & Siegmund, A. (2010). Informationsgewinnung mit Hilfe Geographischer Informationssysteme – Schlüsselkompetenz einer modernen Geokommunikation. *Geographie und Ihre Didaktik, 38*(2), 102-108.

Wallace, J. D., & Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching, 27*(10), 1033-1052.

Wanjohi, K. (2002a). Cultural tourism: a trade-off between cultural values and economic values. In J. S. Akama & P. Sterry (Eds.), *Cultural tourism in Africa: strategies for the new millennium, Proceedings of the ATLAS Africa International Conference, December 2000, Mombasa, Kenya* (pp. 77-93). Arnhem: ATLAS.

Wanjohi, K. (2002b). Perceived socio-cultural impacts of tourism: the case of Malindi, Kenya. In J. S. Akama & P. Sterry (Eds.), *Cultural tourism in Africa: strategies for the new millennium, Proceedings of the ATLAS Africa International Conference, December 2000, Mombasa, Kenya* (pp. 109-124). Arnhem: ATLAS.

Wanner, S., & Kerski, J. J. (1999). *The effectiveness of GIS in High School Education.* Paper presented at the ESRI International User Conference.

Weaver, D. B. (1999). Magnitude of ecotourism in Costa Rica and Kenya. *Annals of Tourism Research, 26*(4), 792-816.

Weidner, W., Bönig, W., Engelmann, D., Gaffga, P., Damian-Guckert, M., Hupke, K.-D., et al. (2006). *Diercke GWG für Gymnasien in Baden-Württemberg 3*. Braunschweig: Westermann.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessung in Schulen* (pp. 17-31). Weinheim: Beltz.

Wendorf, C. A. (2004). Analysis of Covariance (ANCOVA): Overview of the Analysis of Covariance. Retrieved 2011-05-10, from http://www.uwsp.edu/psych/cw/statmanual/ancova overview.html

West, B. (2003). Student Attitudes and the Impact of GIS on Thinking Skills and Motivation. *Journal of Geography, 102*(6), 267-274.

Westermann (Ed.). (2008). *Diercke Weltatlas, Druck A2*. Braunschweig: Westermann.

Wetzel, M. (2014). Gymnasium statt Realschule. Eltern hören seltener auf die Grundschullehrer. *Stuttgarter Nachrichten,* (28.01.2014). Retrieved from http://www.stuttgarter-nachrichten.de/inhalt.gymnasium-statt-realschule-eltern-hoeren-seltener-auf-die-grundschullehrer.f58ead33-4dc4-4088-89a1-bd938c4282b4.html

Wheeler, C. (2009). Seen and Heard at the 2009 ESRI International User Conference. Retrieved 2010-03-22, from http://www.esri.com/news/arcwatch/0809/seen-and-heard.html

Wiegand, P. (2006). *Learning and Teaching with Maps*. London: Routledge.

Wigglesworth, J. C. (2003). What is the best route? Route-finding strategies of middle school students using GIS. *Journal of Geography*(102), 282-291.

Wilder, A., Brinkerhoff, J. D., & Higgins, T. M. (2003). Geographic Information Technologies + Project-Based Science: A Contextualized Professional Development Approach. *Journal of Geography, 102*(6), 255-266.

Wilhelmi, V. (2006). Nachhaltigkeit und Umwelterziehung, Leitbilder des Geographieunterrichts. *Praxis Geographie*(2), 4-8.

Wilson Smith, T., & Colby, A. (2007). Teaching for Deep Learning. *Clearing House: A journal of educational strategies, issues and ideas, 80*(5), 205-210.

Wimdu GmbH. (2014). Mitarbeiter/in Customer Support Deutsch & Spanisch (2nd Level). Retrieved 2014-06-23, from http://de.indeed.com/cmp/Wimdu-GmbH/jobs/Mitarbeiter-Customer-Support-Deutsch-Spanisch-f5df38944899c661

Winship, J. M. (2004). *Geographic Literacy and World Knowledge Among Undergraduate College Students.* M.sc. Thesis, Virginia Polytechnic Institute and State University.

Wolske, K., & Higgs, A. (2010). Power analysis, statistical significance, & effect size. (Contributors: Michaela Zint). Retrieved 2011-05-05, from http://meera.snre.umich.edu/plan-an-evaluation/plonearticlemultipage.2007-10-30.3630902539/power-analysis-statistical-significance-effect-size

Woolfolk Hoy, A. (2004). *What do teachers need to know about self-efficacy?* presented at the American Educational Research Association, April 15, 2004. Retrieved from http://www.coe.ohio-state.edu/ahoy/What%20do%20teachers%20need.pdf

World Resources Institute, Department of Resource Surveys and Remote Sensing Ministry of Environment and Natural Resources Kenya, Central Bureau of Statistics Ministry of Planning and National Development Kenya, & International Livestock Research Institute (Eds.). (2007). *Nature's Benefits in Kenya: An Atlas of Ecosystems and Human Well-Being*. Washington D.C.

Wright, B. D., & Tennant, A. (1996). Sample Size Again. *Rasch Measurement Transactions, 9*(4), 468. Retrieved from http://www.rasch.org/rmt/rmt94h.htm

WTO. (2006). International Tourist Arrivals. Retrieved 2008-04-06, from www.unwto.org/facts/eng/pdf/historical/ITA_1950_2005.pdf

WTO. (2008). Short-term tourism data 2007. *UNWTO World Tourism Barometer, 6*(1), 3.

Wu, M., & Adams, R. (2007). How well do the data fit the model? In: Applying the Rasch model to psycho-social measurement: A practical approach. Retrieved 2011-11-30, from http://www.edmeasurement.com.au/_docs/RaschMeasurement_Part8.pdf

Yang, E.-m., & Andre, T. (2003). Spatial ability and the impact of visualization/ animation on learning electrochemistry. *International Journal of Science Education, 25*(3), 329-349.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). A comparison of two construct-a-concept map science assessments: created linking phrases and selected linking phrases. *Journal of Research in Science Teaching, 42*(2), 166-184.

Yuda, M., & Itoh, S. (2007). Possibilities To Utilize GIS In ESD – From A Research On GIS For Secondary School Teachers In Japan. In S. Reinfried, Y. Schleicher & A. Rempfler (Eds.), *Geographical Views on Education for Sustainable Development. Proceedings Lucerne-Symposium, Switzerland July 29-31, 2007* (pp. 109-114): Selbstverlag des Hochschulverbandes für Geographie und ihre Didaktik e.V.

Zink, R., & Scheffer, J. (2009). GIS auf neuen Wegen in das Klassenzimmer? - Potenziale und Einsatzmöglichkeiten von GIS-Projekten in den Seminaren der gymnasialen Oberstufe. In T. Jekel, A. Koller & K. Donert (Eds.), *Learning with Geoinformation IV - Lernen mit Geoinformation IV* (pp. 121-129). Heidelberg: Wichmann.

Zürl, B. (2005). „GIS - was ist das?" *Geographie heute, 26*(233), 8-11.

Zuzovsky, R. (2007). Capturing the dynamics that led to the narrowing achievement gap between Hebrew-speaking and Arabic-speaking schools in Israel: Findings from TIMSS 1999 and 2003. In IEA (Ed.), *The second IEA international research conference: proceedings of the IRC-2006. Volume 1: Trends in international mathematics and science study (TIMSS)* (pp. 23-44). Amsterdam.

Zweig, S. (1995). *Nirgendwo in Afrika*. München: Langen Müller.

Zweig, S. (2000). *Karibu heißt willkommen*. München: Langen Müller.

Source for Cover Image: part of a map by Claude Bernou (about 1681) taken from http://en.wikipedia.org/wiki/File:Claude_Bernou_Carte_de_lAmerique_septentrionale.jpg, (2010-03-02), opacity reduced, file size reduced

country classification: not all countries in a category represented but areas delineated would comprise; as no students from Estonia, Latvia and Lithuania participated, they did not have to be uniquely assigned; EU: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, UK http://europa.eu/about-eu/countries/index_en.htm; Balkan States: Albania, Bosnia and Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Serbia http://www.state.gov/p/eur/rt/balkans/; Former Soviet Union: Belarus, Armenia, Azerbaijan, Kazakhstan, Kyrgyzstan, Moldova, Turkmenistan, Tajikistan, Uzbekistan, Georgia, Russia, Ukraine; Estonia, Latvia, Lithuania, http://focus-migration.hwwi.de/Russian-Federation.6337.0.html?&L=1; former *Gastarbeiter* (indentured laborer) countries: Italy, Greece, Spain, Turkey, Morocco, Portugal, Tunisia, Yugoslavia (West Germany), Vietnam, Poland, Mozambique and others (East Germany) http://www.wissen.de/wde/generator/wissen/ressorts/finanzen/wirtschaft/index,page=34 86842.html (all October 2011)

Dictionaries etc. were used in some cases to help in the translations.

EAP/PV abbreviation explanation: http://books.google.de/books?id=-RfRfby19O0C&pg=PA68&lpg=PA68&dq=eap/pv+reliability+posteriori+plausible&source=bl&ots=X3Z42nSxp0&sig=ywU6FXLaXRhxERLdJX2gDtzI5tE&hl=en&sa=X&ei=JmKbUanAMZGO4gSX04HYCg&ved=0CCwQ6AEwAA#v=onepage&q=eap%2Fpv%20reliability%20posteriori%20plausible&f=false(2013-08-25)

10. Appendices

10.1 Rasch models (Conquest)

10.1.1 Basic model

Title;

Format pid _____ responses _____;

Model item;

Estimate ! stderr=full;

Show ! >> _____.shw

Itanal ! >> _____.itn;

10.1.2 Model with export (pretest)

Title;

Format pid _____ responses _____;

Model item;

Export parameters >> _____.dat, reg_coefficients >> _____.dat, designma-

trix >>_____.dat, covariance >> _____.dat;

Estimate ! stderr=full;

Show ! >> _____.shw

Itanal ! >> _____.itn;

10.1.3 Model with import (posttest)

Title;

Format pid _____ responses _____;

Model item;

Import anchor_parameters << _____.dat, anchor_reg_coefficients <<

_____.dat, designmatrix << _____.dat, anchor_covariance << _____.dat;

Estimate ! stderr=full;

Show ! >> _____.shw

Itanal ! >> _____.itn;

10.2 Tourism in Kenya

Kenya has been a destination for individual adventurers since before 1900 (Akama, 1999). Having already 39540 tourist arrivals in 1955, from independence (in 1963) on the country saw an increasing development towards tourism as an important economical factor (Akama, 1999). In 2006, international arrivals amounted to about 1.6 million (Ministry of Tourism and Wildlife, 2007b). Tourism accounted for about 12% of the GDP (rank three, after agriculture and manufacturing), 56.2 billion Kshs (Kenyan Shillings) of foreign exchange earnings (rank three, after tea and horticulture), and 400000 formal and 600000 informal employments (Ministry of Tourism and Wildlife, 2007b). It is seen as a possibility for economic growth and job creation (Akama, 1999), thus combating poverty and unemployment (Akama, 2002). Tourism is seen as a possibility to become less dependent on raw products (Akama, 1997, 1999, 2000). However, tourism is an unstable economic sector in Kenya, for instance due to degradation of infrastructure, political and economic crises, dependency on press coverage, the economic situation and changing tourism preferences (Akama, 1999). Real or imagined political instability, for instance, can lead to shifting the tourism from one destination to another with similar offers or to poor business, including for instance the loss of jobs (Akama, 1999; Erhard, 2003; Marshall, 2004). Tourism financing often comes from abroad, increasing external dependencies (Akama, 1999). Many of the jobs for the locals are not paid well (Akama, 2002; Marshall, 2004; Meyer, 1995). Moreover, many of the goods for the hotels, workers etc. are drawn from Nairobi or abroad (Marshall, 2004; Meyer, 1995). Economic leakage to foreign operators etc. is very high (Akama, 1997, 1999, 2000, 2002; Ibrahim, 2003; Marshall, 2004). The local population, despite bearing the major share of the costs of tourism, only receives few percent of the gains from tourism and many living near the tourist centers receive no monetary gain from tourism at all (Akama, 1997, 2000, 2002; Campbell et al., 2003).

In 2006, the average length of stay was 12.1 days, with a hotel occupancy rate of 45.5% (Ministry of Tourism and Wildlife, 2007b). Major countries of origin for tourists besides Kenya are the United Kingdom, Germany, Italy and other

European countries as well as the United States of America (Akama, 1997; Ministry of Tourism and Wildlife, 2007b).

A simplified definition of tourism is "travel for leisure and all the activities and facilities devoted to such travel" (Ministry of Tourism and Wildlife, 2007a, p. 1). Oftentimes, the exact place where the tourists' wishes are satisfied does not matter (Sturm, 1981). Kenya is an attractive destination for tourism mainly because of its exotic animals, "primitive" cultures and beaches in nice climate. It is catering for wishes such as exoticness, getting tanned, temporary escaping from work and enlarging the capital of one's personal image, as has been taken up by travel brochures, advertisement, books and marketing (Akama, 1997, 1999, 2002; Backes, 2005; Dissen, 1990; Job & Metzler, 2003; Odunga & Folmer, 2004; Sturm, 1981; Trichy, 1978; Wanjohi, 2002a). Consequently, tourists are spatially concentrated in few areas in Kenya (Akama, 1997). Besides location attributes such as a higher wildlife density in some areas, this is also due to government politics, and economical processes such as investor preference for areas that are relatively known, have basic infrastructure and seem to be able to generate profits fast (Akama, 1997). Concentration areas include Nairobi, the coast (especially Mombasa, Malindi and Diani Beach), as well as the Maasai Mara, Amboseli, and Tsavo national parks (Akama, 1997, 1999; Meyer, 1995; Ministry of Tourism and Wildlife, 2007a; Weaver, 1999). Thus, the heavily visited sites face environmental degradation and tourist satiation, while the little visited sites could vanish over time (Akama, 1997; Odunga & Folmer, 2004). Tourism numbers vary seasonally. The rainy season from March to May still allows for beach tourism because of sufficient sunshine hours, but safaris face difficulties such as pathways turning into mud and views on wildlife being obscured through vegetation (Esikuri, 1998; Meyer, 1995; Ministry of Tourism and Wildlife, 2007b). Whether and how much other attractions (besides wildlife viewing, and, to a lesser degree, beach tourism) are consumed by the tourists depends also on tourist characteristics such as socioeconomic status, accompanying family, gender and length of stay (Odunga & Folmer, 2004).

Kenya's area of 582350 km² features a variety of land covers: desert and semi desert in the North West, patches of dry savannah in the West and the South, dry forest along the coast and between Nairobi and Mt. Kenya, patches of

tropical rainforest, thorn bush savannah in the rest of the country, besides irrigated and non-irrigated agricultural areas and settlements (Westermann, 2008, pp. 144-145). About 59 percent of the land area has a moderate to high soil fertility (Mwagore, n.d.). However, depending on the source, only about 7 to 17% of the land area can be used for intensive cultivation without irrigation (Baumann et al., 2004; Esikuri, 1998; Federal Research Division, 2007; Meyer, 1995; Mwagore, n.d.). This land area is home to the majority of the population (Esikuri, 1998; Meyer, 1995). The South West has only a low probability for drought, but the rest of Kenya one that is frequent to very frequent (continually to every five years) (Westermann, 2008, p. 132).

"Southern Kenya and northern Tanzania support the greatest concentration of large mammal species across Africa (Reid et al. 1998) and possibly on earth (Sinclair 1995)" (Campbell et al., 2003, p. 2). With the increase of tourism in Kenya over the course of the 20th century, more and more national parks were established (Marshall, 2004). The over 50 protected areas cover about 8% of the country's total landmass (Ibrahim, 2003; Makonjio Okello & Warui Kiringe, 2004). The protection can lead to an overpopulation of animals. For example, if elephants occur in too great numbers, after initially being able to view them better due to a reduced number of trees, in time the landscape becomes unattractive to the tourists (Esikuri, 1998; Gaumnitz, 1993; Trichy, 1978). Moreover, elephant crop raids supply plenty of problems with local farmers, as protected areas seldom include whole ecosystems (Esikuri, 1998; Makonjio Okello & Warui Kiringe, 2004; Trichy, 1978). The local population is often not recompensed for the loss, enhancing the conflicts caused by exclusion from using the areas that have become "protected" (which often include traditionally important dry season grazing ranges) and an increasing human encroachment with accompanying competition for resources (Akama, 2002; Campbell et al., 2003; Makonjio Okello & Warui Kiringe, 2004; Marshall, 2004). This can also lead to retaliative actions such as wildlife killing (Frank, Maclennan, Hazzah, Bonham, & Hill, 2006; Gaumnitz, 1993; Makonjio Okello & Warui Kiringe, 2004). Tourism in Kenya is largely nature based and thus has a potential for supporting biodiversity conservation and environmental protection (Dissen, 1990; Weaver, 1999). Yet, "[...] heavy usage, inadequate regulation, and poor management of both the in-

frastructure and visitor behaviour" (Ministry of Tourism and Wildlife, 2007a, p. 2) led to a deterioration of the few heavy used facilities, especially since many tourism facilities are near fragile breeding or feeding places to make it easier for the tourists to watch the wildlife (Akama, 1997, 2000; Job & Metzler, 2003). Problems include "[...] pollution of water resources, land degradation and unsustainable use of land, air pollution and noise, solid wastes and littering, sewage pollution, aesthetic pollution and introduction of invasive species, physical impacts arising from infrastructure, destruction of marine ecosystems, trampling of vegetation due to off road driving and hiking, anchoring, destruction of fragile ecosystems due to marine sports, and alteration of ecosystems and animal natural behavior due to intense tourism activities" (Ministry of Tourism and Wildlife, 2007a, p. 2) (cp. also Akama, 1997, 2000; Dissen, 1990; Esikuri, 1998; Job & Metzler, 2003; Makonjio Okello & Warui Kiringe, 2004; Weaver, 1999). Additionally, degradation due to overuse occurs in the less suitable areas in which pastoralists are pushed to go due to exclusion from the parks (Ibrahim, 2003). This all can, in turn, endanger tourism which relies on "on pristine and well stocked wildlife sites" (Makonjio Okello & Warui Kiringe, 2004, p. 66) (cp. Akama, 1997, 2000).

Kenya's about 30 million inhabitants are split up in more than 42 ethnic groups (Akama, 2002; Ministry of Tourism and Wildlife, 2007b). The largest African group are the Kikuyu (22% of total population). Non African minorities include Europeans, Indians, Pakistanis and Arabs (Federal Research Division, 2007). The most famous inhabitants however are the Maasai (Wanjohi, 2002a). The literacy rate of the population is estimated to be 75 to 85%, whereby 2003 only about 23% of the respective age group where enrolled in secondary school (ninth to twelfth grade) (Federal Research Division, 2007).

The Maasai culture has been changing through tourism. "Today, traditional cultures are understood and exploited primarily from an economic point of view" (Wanjohi, 2002a, p. 77). In the process, they often loose authenticity and become fixed and adapted to the tourists' expectations and fantasies of "noble savages" (Akama, 2002; Wanjohi, 2002a). Thus, Maasai businessmen or large scale farmers, although existing, are typically not represented in the discourse or marketing generally conducted by non-Maasai agencies (Akama, 2002). This could lead to Kenya loosing its attractiveness for cultural tourists, which often look for less de-

veloped tourist destinations with "unspoiled" communities whose culture has a relatively large gap to their own (Wanjohi, 2002a). Tourism can also increase use of illegal drugs among the youth, prostitution or begging (Akama, 2002; Wanjohi, 2002a). Only very few of the cultures have been paid attention to by the tourism industries (Akama, 1997, 2002; Ipara, 2002). On the other hand, tourism can help to preserve traditional skills and art (Wanjohi, 2002b).

A lot of information about different aspects of Kenya as well as accompanying maps can be found in World Resources Institute et al. (2007).

10.3 GIS data sources

The layers have been translated, cropped, simplified, and partially changed in their geometries.

1. WRI: "Nature's benefits in Kenya: An Atlas of Ecosystems and Human Well-Being":

- selected rivers, towns, streets, hotels

- population density, district boundaries, elephants 1990 and elephant area, airports, land use, altitude, protected areas

2. DEPHA: train lines

3. ILRI: language groups

4. FAO: lakes

5. CIA World Factbook: selected countries (as information source for some attributes)

6. IGAD: protected areas

7. (Westermann data)

8. selected mountains and the Indian Ocean newly created based on existing data

9. climate zones created based on the climate classification of Siegmund/ Frankenberg, the conversion factors of pLV according to Lauer/ Frankenberg and climate data of the CRU

10.4 Additional sources for the learning materials

The photo transparency in study 5 used photos from www.samsays.com and Ogada Ochala. Additionally, some standard cliparts (e.g. book/ computer) were used.

The materials in study 5 also used e.g. data from the WTO (2006, 2008) and the ThaiIndian Newsportal/ WRI[1] for international tourist arrivals.

The additional materials (Maasai, Wangari Maathai, travel destinations) also used authentic materials, namely

- a very brief text excerpt from www.hauser-exkursionen.de/afrika/kek16ein_platz_fuer_wilde_tiere.html

- a map from www.ogiek.org/faq/maasai.htm

- a short text based on Terra GWG 3/4 and www.historyworld.net/worldhis/ PlainTextHistories.asp?historyid=ad21

- short experts from Diercke Geographie 2/3, pp. 232-233 as well as www.scinexx.de/dossier-detail-225-10.html

- a short text based on www.nobelprize.org and www.gbmna.org as well as a picture of Wangari Maathai from wikimedia commons

- a short text with information from www.gbmna.org

Moreover, based on the wish of a teacher during the study, a solution sheet was created.

---

1 apparently no longer available; slightly different data can be found e.g. in UNWTO (n.d.) for many countries and Ministry of Tourism and Wildlife (2007b) for Kenya